DECISION-LEVEL GAIT FUSION FOR HUMAN

IDENTIFICATION AT A DISTANCE

by

Amer Al-Tayyan

A Thesis Presented to the Faculty of the
American University of Sharjah
College of Engineering
in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science in
Electrical Engineering

Sharjah, United Arab Emirates

June 2014

Approval Signatures

We, the undersigned, approve the Master's Thesis of Amer Al-Tayyan.

Thesis Title: Decision-Level Gait Fusion for Human Identification at a Distance

Signature   **Date of Signature**
                                                                 (dd/mm/yyyy)

_____   _____
Dr. Khaled Assaleh
Professor, Department of Electrical Engineering
Thesis Advisor

_____   _____
Dr. Tamer Shanableh
Associate Professor, Department of Computer Science and Engineering
Thesis Advisor

_____   _____
Dr. Serkan Kiranyaz
Professor, Department of Electrical Engineering
Thesis Committee Member

_____   _____
Dr. Michel Pasquier
Associate Professor, Department of Computer Science and Engineering
Thesis Committee Member

_____   _____
Dr. Mohamed El-Tarhuni
Head, Department of Electrical Engineering

_____   _____
Dr. Hany El Kadi
Associate Dean, College of Engineering

_____   _____
Dr. Leland Blank
Dean, College of Engineering

_____   _____
Dr. Khaled Assaleh
Director of Graduate Studies

Acknowledgements

Abstract

Gait Recognition is one of the latest and most attractive biometric techniques currently under research, due to its potential application in identification of individuals at a distance, unobtrusively and even using low resolution images. In this thesis a comprehensive study of the gait problem is presented, covering gait databases and the different approaches used for preprocessing, feature extraction and classification. The objective is to achieve a robust technique that performs well, independently of the input data and the many covariates that affect this behavioral biometric technique. Firstly, gait data is processed using three gait representation methods as the features sources; Accumulated Prediction Image (API) and two novel gait representations namely; Accumulated Flow Image (AFI) and Edge-Masked Active Energy Image (EMAEI). Secondly, each of these methods is tested using three matching schemes; Image Projection with Linear Discriminant Functions (LDF), Multilinear Principal Component Analysis (MPCA) with K Nearest Neighbor (KNN) classifier and the third method: MPCA+ Linear Discriminant Analysis (MPCALDA) with KNN classifier. Gait samples are fed into the MPCA and MPCALDA algorithms using a novel tensor-based form of the above-mentioned gait representations. We end-up having nine recognition modules which are analyzed individually using four different experimental setups and compared to the results reported in six of the most recent papers that used the same database and the same experimental setups. Finally, decisions from the nine recognizers are fused using decision-level (majority voting) scheme. A comparison between unweighted and weighted voting schemes for final decision is also shown. The experimental results show clearly that the proposed approach outperforms the state-of-the-art gait approaches used in the literature, and reports the highest recognition rates known to the date of writing this report. As a result of the comprehensive study and extensive experiments, we conclude that model-free gait approaches, particularly spatio-temporal and energy-based methods, are the best choice in a gait recognition system used for human identification. We also note that single classifiers may not be reliable and robust to deal with gait recognition, and a fusion scheme that combines the power of each of the base classifiers is needed.

Search Terms: *Gait Recognition, Human Identification, Gait Tensors, Multilinear Subspace Learning (MSL), Decision-Level Fusion*

Table of Contents

List of Figures

List of Tables

Abbreviations

AEI	Active Energy Image
AFI	Accumulated Flow Image
AFT	Accumulated Flow Tensor
API	Accumulated Prediction Image
APT	Accumulated Prediction Tensor
BTA	Baysiean Tensor Analysis
CASIA	Chinese Academy of Science and Institute of Automation
CCD	Charged Coupled Device
CDA	Canonical Discriminant Analysis
CGI	Chrono Gait Image
CMU	Carnegie Mellon University
CSA	Concurrent Subspace Analysis
DATER	Discriminant Analysis with Tensor Representation
DARPA	Defense Advanced Research Projects Agency
DCT	Discrete Cosine Transform
DTW	Dynamic Time Warping
EMAEI	Edge-Masked Active Energy Image
EMAET	Edge-Masked Active Energy Tensor
EMP	Elementary Multilinear Projection
FAR	False Accept Rate
FIR	Finite Impulse Response
FPT	Full Projection Truncation
FRR	False Reject Rate
GEI	Gait Energy Image
GFI	Gait Flow Image
GHI	Gait History Image
GLRAM	Generalized Low-Rank Approximation of Matrices
GPCA	Generalized Principal Component Analysis
GPPE	Gait Pal and Pal Entropy
GSP	Gait Structural Profile

GTDA	General Tensor Discriminant Analysis
HMM	Hidden Markov Model
HUMABO	Human Monitoring and Authentication using Biodynamic Indicators
ITA	Incremental Tensor Analysis
KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
LDF	Linear Discriminant Functions
MAEI	Masked Active Energy Image
MoBo	Motion of Body
MPCA	Multilinear Principal Component Analysis
MPCALDA	MPCA with Linear Discriminant Analysis
MSL	Multilinear Subspace Learning
NIST	National Institute of Standards and Technology
NMPCA	Non-negative MPCA
PMS	Procrustes Mean Shape
PCA	Principal Component Analysis
PSA	Procrustes Shape Analysis
RMPCA	Robust MPCA
SAD	Sum of Absolute Difference
SEI	Shifted Energy Image
SVM	Support Vector Machines
2DLDA	Two-Dimensional Linear Discriminant Analysis
2DLPP	Two-Dimensional Locality Preserving Projection
2DPCA	Two-Dimensional Principal Component Analysis
TROD	Tensor Rank-One Decomposition
TR1DA	Tensor Rank-One Discriminant Analysis
TTP	Tensor-to-Tensor Projection
TVP	Tensor-to-Vector Projection
UCSD	University of California, San Diego
UMLDA	Uncorrelated Multilinear LDA
UMPCA	Uncorrelated MPCA
USF	University of South Florida

UWV	Unweighted Voting
VVP	Vector-to-Vector Projection
WV	Weighted Voting

Chapter 1: Introduction

Gait is defined as "the *coordinated* and *cyclic* combination of movements that result in human *locomotion*" [1]. As such, and based on these repeated patterns, it can be used to identify people. The use of gait traits in biometrics is increasingly attracting researchers' and scientists' interest. This is mainly due to its potential in identification of individuals at a distance. Since 1994 (marked by the work of Niyogi and Adelson [2]) until this moment in time, many approaches were introduced trying to analyze, understand and extract gait features, and hoping to overcome the many challenges that impact gait recognition, and to improve the overall systems' performance. The major challenging factors would be; walking style and speed, time span, walking surface, viewpoint and clothing. Gait analysis methods might be classified into Model-Free and Model-Based; referring to the approach of dealing with the body statics and/or dynamics. One important decision to make before testing any gait recognition algorithm is the selection of the database. The database should be suitable to the application of study, and will be used to train, test and evaluate the system. Approaches that aim to design a system for surveillance human monitoring, for example, may use image sequences captured outdoor, and considering different conditions, in order to test against real life scenarios. Different approaches can be compared using different databases. The ultimate goal of any gait recognition system is to achieve high recognition rates and being robust under all conditions.

1.1 Motivations and Challenges

Due to the potential of gait and its promising applications, extensive researches continue to highlight and examine the main problems and challenges that hinder the commercialization of these systems. Although high recognition rates were achieved in most of those studies, it was resulted from limited number of subjects, limited covariates and highly controlled environments. In reality, conditions are much more challenging, a matter that necessitates the development of a gait recognition system that is immune to noise and security attacks, and almost independent of those different covariates. The application of gait in biometrics requires specific focus on

some covariates like the view angle, clothing, footwear, time span and carrying conditions.

Therefore, this study aims to go a step further on the path to achieve a robust gait recognition system that is acceptable and workable worldwide; for different cultures, dress codes, walking styles and conditions. We examine thoroughly the different gait approaches and databases used in the literature and their performance and applicability.

1.2   Contribution

The main contributions in our work include:

a. A comprehensive and extensive study of the gait problem introducing literature review, tests and analysis of the major and state-of-the-art reported techniques. This should provide a good reference and basis for many similar studies in the future. Our study has affirmed the assumption that model-free, especially spatio-temporal (accumulated error) and energy-based, methods are more effective gait methods. Besides, we conclude that decision-level fusion from multiple gait recognition methods performs much better than each of the single methods involved.

b. Exhaustive experiments on, and optimization of, the Accumulated Prediction Image (API) method from [3], section (2.2).

c. Proposing two novel gait representations; the Accumulated Flow Image (AFI) and Edge-Masked Active Energy Image (EMAEI), found in sections (2.3) and (2.5) respectively.

d. Further implementation and testing of tensor-based feature extraction methods for gait recognition. The first implementation of these methods on the gait problem was by Haiping et al. [4] using the Multilinear Principal Component Analysis (MPCA) method. However, we have implemented the method for the first time on the CASIA B dataset (described in 2.1.1), introducing novel tensorial gait representations that proved to be competitive.

e. Testing the outcome of fusion on the overall performance of gait recognition systems, especially decision-level fusion.

f. Achieving the highest recognition rates in the categories, as to the date of writing this report, by comparing our results to several recently-published papers.

1.3 Background

1.3.1 Biometrics

Biometrics is the science and techniques that deal with the *automatic measurement* of *physiological* or *behavioral* characteristics of humans for the purpose of *identification* or *verification*.

Compared to other identification credentials like passwords or cards, the latter can be forgotten, lost or stolen. While biometric data is inherent in humans and is, hence, more secure. For data to be considered biometric, it has to be unique among individuals, universal, permanent, collectable (measurable), acceptable and accurate.

Examples of physical biometric include fingerprint, palm print, iris scan, retina scan, face scan, hand geometry, vein pattern, ear shape and body odor. Behavioral biometric include speaker, signature dynamics, keystroke dynamics, gait and Human-Computer Interaction.

The difference between identification and verification is that in the first, one person's template is compared to all others in the database to identify him/her. It is therefore known as *1:N* process. In verification, the person claims his/her identity, before the system verifies it by comparing it to the stored template. This mode is known as *1:1* process.

Biometrics, however, is not perfect. Therefore, there is some performance metrics used to evaluate the system. The most important parameters are the False Accept Rate (FAR) and the False Reject Rate (FRR). FAR may occur when more than one person have similar biometric data. It can be reduced by increasing the threshold level. However this would also increase the FRR. This latter means a valid person is rejected, and this is due to bad correlation between the enrollment and verification templates. It can be solved by re-enrollment and by controlling the environment. There will always be trade-off between FAR and FRR.

1.3.2 Gait Recognition

Automatic recognition by gait is attractive compared to other biometric techniques. Its major potential as a biometric technique is in identification of humans at a distance [5]. It can be applied to subjects unobtrusively and even without their cooperation or awareness. Compared to face recognition, for example, which is also useful in surveillance applications, the latter almost fails for outdoor recognition over a distance. Here, gait does not require high resolution images. Different approaches were presented and using different databases. Identification rates have exceeded 90% under similar conditions for the training and the testing databases [6]. When taken outdoor, other challenges come into consideration, primarily the changing environment, the change in walking style/speed and illumination. Performance analysis against the different covariates is the topic of most of the research work nowadays.

In [1], Little and Boyd defined gait as "the *coordinated* and *cyclic* combination of movements that result in human *locomotion*". The keywords in this definition; coordinated, cyclic and locomotion are what distinguish gait from other forms of movements. And these movements should produce unique spatio-temporal patterns that may be used to identify people in a biometric system.

It was shown in different fields (such as psychology, biomechanics and medicine) that gait signature is unique and can be used to identify people. As we know, the most important condition to consider data in a recognition system, biometric, is uniqueness. Literature in biomechanics and medicine show that every person performs a unique and repeatable pattern in walking. This pattern is usually measured over one full stride. Figure 1 shows a typical gait (or walking) cycle.

Johansson proved with experiments in [7] that human is able to identify a person via a moving light dots pattern. When these dots were static, it was not easy to recognize a human pattern. However, this proves the human perception of gait.

Murray et al. [8] showed that movement of the upper part (thorax and pelvic) of the body is subject to variance. Therefore, it is the lower part (limb, hip, thigh and heel) which is primarily used in gait analysis.
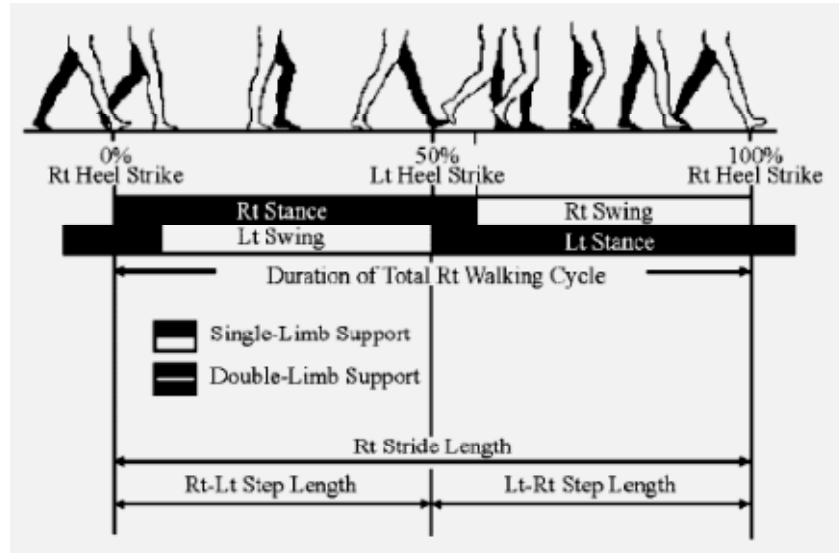
Figure 1: A typical gait cycle [5]

Factors that might affect gait or the perception of gait include; skeletal dimensions [9], terrain/surface [10, 11], injury [8], footwear [11, 12], muscle development [1], fatigue [1], viewpoint [11], training (athletic or military) [1], time [11], culture [1], carrying condition [13-16] and clothing [13-16]

Among the few studies that examined the robustness issue of gait, major contribution was by Davrondzhon Gafurov, as in [17] and with his team in [18]. They studied the scenarios of both passive (unknowledgeable) imposter, and active (knowledgeable) imposter. While it was assumed that gait is robust to most spoof attacks, the study in [17] proved that imposters who know the gender of some persons in the database may cause a threat. Also the study in [18] proved that knowing the person of the closest gait in the database might be risky. Therefore, the issue should be addressed and considered in the design process. A possible example of counter-measure can be implementing fusion techniques.

Gait robustness was also the topic for some studies, like [19] and [20]. In [19], Lee et al. used dimensionality reduction techniques to extract the spatio-temporal features of gait.

1.3.3 Gait Databases

Just like any other pattern recognition system, gait analysis requires the use of pre-acquired samples of data for training and testing. In the case of gait, these samples are video sequences of different subjects, walking. For robustness purpose, it is

necessary to have these data unique and practical. This can be accomplished by capturing as many subjects as possible in multiple sequences; from the two genders, under different conditions and over different time periods. When such database is used, we anticipate having more accurate evaluation of the system and more generalized results.

1.3.4 Tensor-based Data Representation

Tensors are multi-dimensional objects, or *n*-mode objects, where *n* is the number of indices in the tensor representation. The fact that all images and video sequences used in biometric techniques, such as face recognition and gait recognition, use tensors as their data, makes it intuitive to try to extract features from the data in its native tensorial form instead of reshaping it into a vector form, or even matrix form. Face templates (grayscale images), for instance, are 2-mode tensors. Whereas grayscale video sequences used in gait recognition are 3-mode tensors: the two spatial dimensions and the temporal dimension. When adding the number of samples (subjects), the problem becomes of 4-mode tensors.

On the other hand, it is also common in such biometric applications to have high dimensional feature space. And in pattern recognition, this may lead to the so-called curse of dimensionality, especially when dealing with small number of samples. It becomes useful in similar conditions to implement the dimensionality reduction techniques prior to feature extraction and classification, so as to map high-dimensional space into a lower-dimensional space while trying to keep the correlated data, and reduce redundancy. And if dimensionality reduction techniques are to be used with multi-mode objects, it becomes intuitive to keep them in their tensorial representation. This is compared to reshaping data into vector form, as is the case in linear techniques like PCA and LDA, which leads to changing the original data structure. Using tensors means dealing with multi-mode data in its native form, and thus hoping to produce more discriminative and correlated features.

1.3.4.1 Basic Multilinear Algebra

An *N*th-order tensor can be represented as $\mathcal{A} \in R^{I_1 \times I_2 \times \dots \times I_N}$, where N is the number of indices (modes), and each I_n represents the *n*-th mode of $\mathcal{A}$. Figure 2

shows an example tensorial representation of a tensor object $\mathcal{A}$ and its vectorial components.

<div align="center">(a)   (b)   (c)   (d)</div>
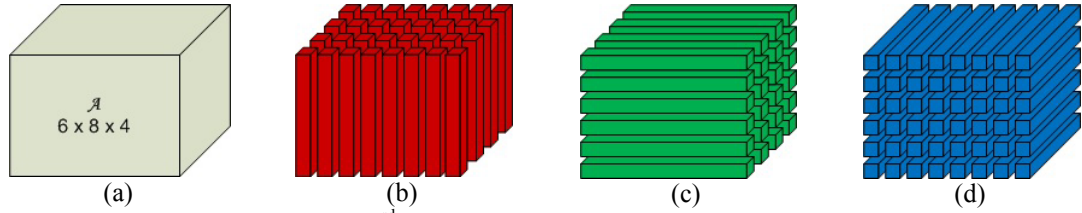
Figure 2: Illustration of 3rd order tensor and its vectors; (a) tensor object, (b) its first-mode vectors, (c) the second-mode vectors and (d) the third-mode vectors.

The n-mode vectors above are the I_n dimensional vectors obtained by varying its index (same mode) while keeping the other indices fixed. This way, it is possible to unfold the tensor into a single-order vector.

1.3.4.2 Multilinear Projections

Multilinear projections deal with mapping tensors into lower dimensional subspaces. In general, there are 3 types of multilinear projections, based on the shape of input and output:

1. Vector-To-Vector Projection (VVP): this is similar to the conventional linear projection as used in PCA and LDA. It means to map a vector x (of dimension I) into another vector y (of dimension P), using projection matrix U (of dimension I x P, where $P \ll I$). Or $y = U^T x$

   Therefore, when the input to VVP is tensor, it should be vectorized first before projection.

2. Tensor-to-Vector Projection (TVP):

   This method is equivalent to multiple projections from a tensor into a scalar, or an approach called Elementary Multilinear Projection (EMP), which in other words means that converting a tensor $\mathcal{X}$ into a vector y of dimension P is accomplished using P EMPs. The detailed description of the EMP is out of the scope of this study.

3. Tensor-to-Tensor Projection (TTP):

   It means to project an Nth-order tensor $\mathcal{X}$ (of n-mode dimensions $I_1, I_2... I_N$) into another tensor $\mathcal{Y}$ at a lower tensorial space (of n-mode dimensions $P_1, P_2... P_N$) using N projection matrices, $U^n \in R^{I_n \times P_n}$. It is accomplished in N

steps by mapping the *n*-mode vector every time or in one step using 1-mode projection matrix that would project each 1-mode vector into a lower dimensional vector. This is the method implemented in the MPCA algorithm, as we shall see in section (3.2)

1.4   Literature Review

1.4.1 Gait Databases

Table 1 summarizes the databases used in the recent studies on gait recognition sorted by date. And here are briefs of the major ones:

In UCSD database [21-23], six subjects were imaged in the University of California, San Diego. Every subject was required to walk in a circular path that passes through the field of view of the camera before a fixed background. Seven image sequences were taken for each subject, thus a total of 42 sequences overall.

Table 1: Summary of major databases used in gait analysis and experiments

Database	Size		Scene/Background	Covariates
	Subjects	**Sequences**		
UCSD, 1998	6	42	Wall Background	Time (minutes)
CMU MoBo, 2001	25	600	Indoor, Treadmill	View angle, Walking Speed, Surface Incline, Carrying Conditions
Maryland, 2001	25	100	Outdoor, ground, 30m away	View angle, Time
	55	222	Outdoor, top-mounted	View angle, Time
MIT, 2001	24	194	Indoor, floor	Time (3 months)
Southampton, 2001 (New SOTON)	28	112	Indoor floor (green background)	Time, footwear
CASIA A, 2001	20	240	Outdoor, three views (0°, 45°, 90°)	View angle
NIST/USF, or Gait Challenge, 2002	122	1870	Outdoor	View angle, Surface, Shoe, Carrying Condition, Time (6 months).
CASIA B, 2005	124	13640	Indoor, multi-view	View angle, Clothing, Carrying Condition
CASIA C, 2005	153	612	Outdoor, Infra-red (night)	Walking conditions
HUMABO, 2007	75 51	- -	Indoor, several times, stereoscopic cameras	View angle
ACTIBIO, 2007	28	-	Indoor, increasing complexity	View angle, walking style
AUS, 2009	103	1030	Indoor, 0° view, right-to-left and left-to-right	Clothing

Motion of Body (or MoBo) database [21, 24-27] was collected in the labs of Carnegie Mellon University (CMU). It contains 600 image sequences from 25 subjects (23 males, 2 females). Each subject was imaged using 6 x 3CCD cameras simultaneously, recording at 30 fps, and repeated with four different types of walking; slow walk, fast walk, inclined walk and slow walk holding a ball. Figure 3 shows samples of this database.

Figure 3: Sample from CMU MoBo Dataset with 6 different viewpoints [26]

Maryland database [21, 24-26] contains two datasets of people walking outdoor. In the first dataset, the camera was mounted horizontally against the subject. In the second dataset (as seen in figure 4) two orthogonal surveillance cameras were mounted on top of the building. 55 subjects (46 males, 9 females) were imaged walking in a T-pattern with four different poses: frontal, right, left and back. This database is good to test for actual surveillance examples. Figure 4 shows samples of this database.

Figure 4: Sample from Maryland Data Set [26]

MIT database [26] consists of 194 sequences, resulted from imaging 24 subjects twice or four times, perpendicular to a single camera view. Each subject was asked to repeat the walking twice, in two 3-month separated sessions.

Southampton database [22, 26] consists of 112 video sequences resulted from imaging 28 subjects, with four sequences each. People were asked to walk on a track perpendicular to the camera view, and against a green background, resulting in clean silhouettes. Figure 5 shows samples of this database.

Figure 5: Sample from Southampton Data Set [26]

CASIA Dataset A [28, 29] was developed by the Chinese Academy of Science and Institute of Automation. It consists of 20 subjects; each was imaged in 12 sequences. These are 4 sequences in each of the three directions; 0°, 45° and 90°. This Dataset includes 19139 images. Figure 6 shows three samples.

Figure 6: Samples of the CASIA Dataset A [28]

One of the largest gait databases is the NIST/USF, or Gait Challenge database. This large database was collected as part of the Gait Challenge Project [11] that was funded by the DARPA (Defense Advanced Research Projects Agency) Human Identification at a Distance (HumanID) Program. It consists of 1,870 video sequences from 122 subjects, and covers five different covariates; view point, footwear, walking surface, carrying condition and time between sequences. The database has been one of the mostly used in gait recognition experiments, as in [24, 25, 27, 30-33]. Figure 7 shows four samples.
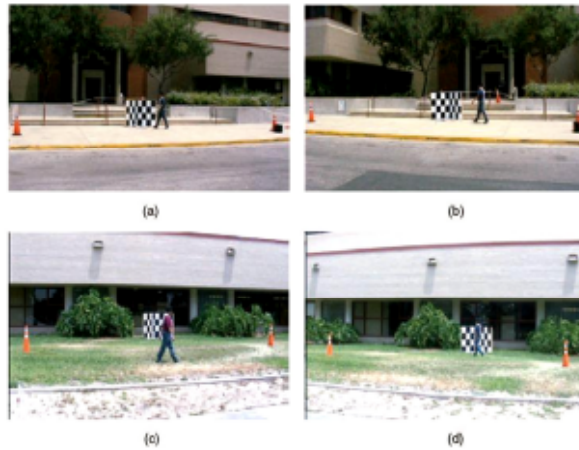
Figure 7: Samples of the NIST/USF Dataset; a) left view walking on concrete b) right view on concrete c) left view on grass, and d) right view on grass. [11]

CASIA Dataset B is a large multi-view dataset that includes 124 subjects, each was imaged in 11 different views and three different conditions; view point, clothing (coat) and carrying condition (bag). This resulted in 13640 image sequences. Figure 8 shows three samples of this dataset. This dataset is being extensively used in gait experiments as in [13-16, 34-40].

| (a) | (b) | (c) |
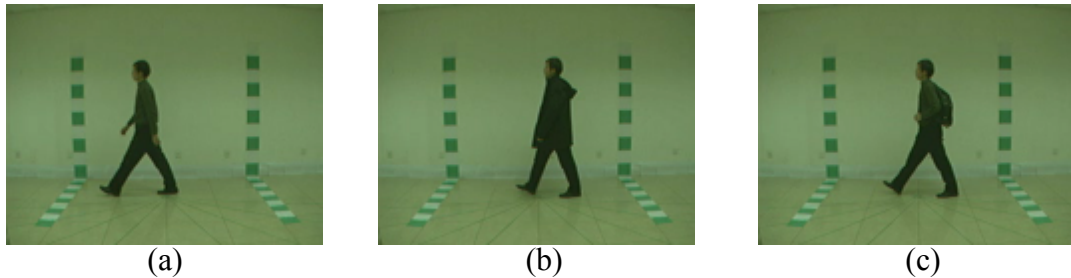
Figure 8: Three samples of CASIA B Dataset, (a) normal condition, (b) carrying bag, and (c) wearing coat

Another dataset by CASIA is Dataset C [13]. It consists of 153 subjects, each was imaged in 4 sequences under four different walking conditions; slow walking, normal walking, fast walking and normal walking with a bag. These sequences were captured at night time using infra-red camera.

HUMABO (Human Monitoring and Authentication using Biodynamic Indicators and Behavioral Analysis) Database [41, 42] is a proprietary database of gait sequences captured indoor, on a fixed path. Subjects were asked to walk normally several times on the same path, and their fronto-parallel walk captured using

stereoscopic cameras. The database is composed of two sessions with 6-month gap; the first one consists of 75 subjects, and the second session consists of 51 subjects.

ACTIBIO [41] is another proprietary indoor database developed also in two sessions with 2-month gap, and 28 subjects in each. The subjects were asked to walk several times with increasing complexity.

In 2009, video sequences from different subjects were collected at the American University of Sharjah (AUS). Undergraduate students aged 18-22 were asked to walk normally for about 10m, being 10m away from a video camera. Each subject had to walk from right to left and backwards for an average of five times. The dataset was mainly collected to study the effect of the gulf costume on gait recognition. 20 males and 33 females in local Gulf costume, as well as 39 males and 11 females in casual costume have participated. AUS Database was used successfully in [3]. Figure 9 shows four samples of the AUS database.

(a)	(b)
(c)	(d)

Figure 9: Samples of the AUS Database: (a) Male in gulf costume (b) Female in gulf costume (c) Female in casual costume (d) Male in casual costume.

1.4.2 Gait Approaches

One of the earliest approaches for automatic gait recognition is dated back to 1994, and introduced by Niyogi and Adelson [2] who used spatiotemporal (XYT) patterns of gait. Later, Sarkar et al. [11] have developed a large database for gait as well as new techniques, based mainly on silhouette extracting. Their work was part of the DARPA HumanID at a Distance Program.

The survey in [43] has categorized gait approaches based on the technology used into; Machine Vision-Based, Floor Sensor-Based and Wearable Sensor-Based. However, most of other researchers, as in [24, 41, 44], categorized these approaches into: Model-Free (or Feature-Based [41]) and Model-Based. The model-free approaches mostly utilize the spatial or spatiotemporal characteristics of the subject, or silhouette shape, and later motion of this shape. And they deal with extracting statistical features from the gait image sequence. Model-based methods use the static and/or dynamic features of human body, like stride length, stride speed and cadence [45] to create model of the human body. In spite of being more reliable and immune to noise, model-based methods have high computational complexity which make them less desirable compared to model-free methods. Table 2 sorts the major approaches chronologically. And here is a summary of them:

1.4.2.1 Model-Based Approaches

Instead of measuring the shape (and probably the motion) of the body's silhouette, these methods create models and/or structures for the body, based on the different parts' movements (torso, leg, stride and cadence-frequency).

Lee and Grimson [46] modeled the human body into seven regions, and represented each region by an ellipse. Then features of each region are extracted. Cunado et al. [47] used models of the legs, as they proved harmonics of the legs motion. This model was used, then, for recognition. Yoo et al. [48] modeled the body's regions into sticks. And Yam et al. [49] used the pendulum-like movement of the legs as a model. Dockstader et al. [50] used a model of thick lines connected with points for different parts of the body and benefited from the pendulum motion of the lower part for feature extraction. Bobick et al. [51] and BenAbdelkader et al. [52] have adopted the structural model of stride parameters to extract gait features.

Joint Trajectory Patterns method models the joint angles' movements and uses them as the gait signature. Tanawongsuwan et al. [53] and Wang et al. [45] applied this method to structure different joints trajectories in the upper as well as the lower parts of the body, including the torso, thighs, legs and feet.

Zhang et al. [27] used a blend of model and structure to describe statics and dynamics of the gait in five-link biped model. Features were extracted from the upper

and lower body parts, described by Fourier Analysis and classified by HMM. They used the NIST/USF and MoBo databases for testing,

Fortuny-Guasch et al. [54] extended the application of radar transceivers for the automatic target recognition, which is in this case based on the human gait. Their approach is evolved from the fact that every person emits a unique Doppler frequency during motion. Their system consists of three transceivers that act simultaneously to map the Doppler signals from the body. First, the body was modeled into parts, and then frequency contribution of each part was determined leading to a radar model of the human body (Doppler Signature).

Table 2: List of Major Gait Approaches

Year	Approach	Database(s)
1998	Shape of Motion [23]	UCSD
2001	Joint-angle Trajectories [53]	Local with sensor
2002	Key Frames [26]	MoBo, Southampton, UMD & MIT
2003	Statistical Relationships [30]	NIST-USF
	Eigenspace Transformation [56, 57]	CASIA A
	Symmetry Analysis in Silhouettes [22]	UCSD
2004	Five-link biped model [27]	USF, MoBo
	Average Silhouette [55, 31]	NIST-USF
2005	Self-Similarity [21]	UCSD, MoBo, UMD
	Gait Period & Silhouette Similarity [11]	NIST-USF
2006	Gait Energy Image (GEI) [31, 33]	USF
2007	Shape Variation-Based Frieze Pattern [19]	MoBo, USF
	Motion Silhouette Contour Templates (MSCTs) and Static Silhouette Templates (SSTs) [62]	SOTON, USF
2008	Fisher Discriminant Analysis [58]	CASIA A
2009	Accumulated Prediction (Difference) [3]	AUS
2010	Soft Biometrics [41]	HUMABIO, ACTIBIO
	Motion Contour Image (MCI) [29]	CASIA A
	Active Energy Image (AEI) [13]	CASIA B, C
2011	Gait Flow Image (GFI) [63]	USF
	Multimodal Face-Gait Fusion [64]	Human Actions
	View Transformation Model (VTM) [37]	CASIA B
2012	Procrustes Shape Analysis (PSA) and Elliptic Fourier Descriptors (EFD) [65]	MoBo, USF
	Shifted Energy Image (SEI) and Gait Structural Profile (GSP) [14]	CASIA B, MoBo, ACTIBIO
	Chrono-Gait Image (CGI) [16]	USF, CASIA B, SOTON
	Metric Learning [38]	CASIA B
2013	Finite Impulse Response (FIR) for motion of legs. [40]	CASIA B
	Gait Pal and Pal Entropy (GPPE) Image [66]	CASIA A, B, C
	Feature Subset Selection [34]	CASIA B
	Interval Valued Features [15]	CASIA B
	Discriminative Set matching [36]	CASIA B
	Procrustes Mean Shape (PMS) [39]	CASIA B, SOTON
2014	Time-sliced Averaged Motion History Image (TAMHI) [67]	CASIA B, OU-ISIR, MoBo

1.4.2.2 Model-Free Approaches

Spatio-Temporal Analysis method studies 3D (XYT) representation of gait images, including gray-scale images, optical-flow images and binary silhouette images. Among the earliest who adopted this method were Niyogi and Adelson [2], considering first XT dimension, and then extending it to XYT. In this method, patterns of the head and legs were of interest, being represented by five-stick models, before features were extracted.

Little and Boyd introduced in [23] the Optical Flow (Shape of Motion) concept. It deals with the distribution of flow in images, represented by the shape of motion. Phase features were extracted and used to evaluate the system. The Nearest Neighbor method was used for classification.

Sarkar et al. [11] used the Gait Period and Silhouette Similarity to segment the subject of interest from its background. The resulted object is called Silhouette, and is been used to study gait in many papers that followed. They conducted 12 experiments of increasing difficulties using NIST/USF Dataset. They developed the Baseline Algorithm which extracts the silhouettes features using temporal correlation, computes the gait period to know the spatial-temporal correlation and computes similarity between subsequences by normalization, using the median for similarity measure. The study concluded that time span and surface type have the largest impact on gait quality and analysis, viewpoint and carrying conditions have medium impact and footwear has a small impact.

Quasi Gait Recognition is achieved by measuring skeletal dimensions as in the work of Bhanu and Han [9]. Researches continue in this direction to prove that skeletal dimensions of human beings really control and affect their gait style.

Collins et al. [26] compared the key frames in a silhouette sequence with the training frame. Features were extracted using normalized correlation. Experiments used the MoBo, UMD, Southampton and MIT datasets. Classification was done via Nearest Neighbor method.

Vega and Sarker [30] implemented statistical relationships to image features using motion model. NIST/USF database was used to evaluate the results.

Kale et al. [24] considered feature vectors for the width of the outer contour of the silhouette. They used MoBo, UMD and NIST/USF databases for evaluation and Hidden Markov Models for classification. Sundaresan et al. [25] did a similar work,

yet considering the feature vector of the whole silhouette. They used the USF for experiments.

BenAbdelkader et al. [21] implemented self-similarity between silhouettes to extract features similar to the way used in eigenfaces for face recognition, and the Principal Component Analysis (PCA). Classification is done using KNN classifier. In this method database of UCSD was used, and extended to MoBo and UMD databases. The method is sensitive to noise and light.

The Average Silhouette method appears in the research papers of Liu et al. [55] and Han et al. [31]. It is claimed to be the simplest method of silhouette feature extraction, by averaging silhouettes of the same subject. It omits some of the detailed features, based on selecting common features. This would simplify the algorithm, however reduce the overall performance.

Murase and Sakai [56] used eigenvalue decomposition for feature extraction of gaits in eigenspace. Huang et al. [57] combined eigenspace Transformation to Canonical Space Transformation. Wang et al. [28] also implemented eigensapce Transformation to get gait features. They used CASIA A dataset for their experiments, implementing KNN classifier (Exemplar Method).

In [19] Lee et al. capture gait information over time and extract features from rows and columns of normalized frame difference images. In their method, called Shape Variation-Based Frieze Pattern, they tested the algorithm on MoBo and NIST/USF databases.

Su et al. [58] used the Kernel Fisher Discriminant Analysis classification method for the recognition of the gait features. Features were obtained from the periodic sequence width images, capturing both static and dynamic features. They used CASIA A dataset for training and testing.

Shanableh et al. [3] have applied the Prediction Error computation technique, used in digital video coding, to the gait problem. Each frame was subtracted from the previous as well as the next one, resulting in accumulated prediction image. The operation was repeated for even frames and odd frames separately, resulting in two Accumulated Prediction (AP) Images per subject. These images were, then, used to extract features by Radon Transformation. Experiments were done on the AUS database, using Polynomial Network for classification.

Moustakas et al. [41] used soft biometric features taken from geometric gait features of the body height and stride. These features were determined using Radon Transforms and Gait Energy Image techniques. The soft biometric features were, then, fused, and probabilistic methods were used for extraction. For evaluation purpose, they used HUMABIO and ACTIBIO databases. The research resulted in enhanced CMS (for identification) and FAR-FRR (for verification) curves.

Han et al. [31, 33] used Gait Energy Image (GEI) that explains dynamics and statics of body features using image sequences. They used NIST/USF database for evaluation.

Other Methods include:

- Measuring Area of Silhouettes as in [59].
- Symmetry Analysis in Silhouettes: Hayfron-Acquah et al. [22] produced symmetry maps for humans using Sobel operators. They proved to be insensitive to noise. The UCSD dataset was used for experiments with a k-nearest neighbor classifier (k=1).
- Shutler et al. [60] used moments of silhouettes.
- Bhanu et al. [32] approach used kinematic as well as stationary features, with the help of a 3D model of the silhouettes, and utilizing the shape and structure of the silhouettes, testing with NIST/USF database.
- Liu and Zheng [61] introduced the Gait History Image (GHI) approach.
- Huang et al. [14] used feature-level fusion of Shifted Energy Image (SEI) and Gait Structural Profile (GSP) in their experiments.

1.4.3 Fusion Schemes

Recent researches in biometrics have shown that single-method (or single-modality) biometric systems can be susceptible to noise, sensors' sensitivity and redundancy of features [68, 69]. This is particularly an issue in behavioral techniques, such as gait. Therefore, the typical solution would be using multimodal biometric techniques, or fusing information from multiple sensors, feature extractors or classifiers. In other words, fusion at feature-level, score-level or decision-level. Fusion may also take place at the sensor level (before feature extraction). An example of this is the fusion of 2D and 3D faces using two different sensors. Fusion techniques

bring to biometrics advantages like higher recognition rates, overcoming the issue of small sample size or training data, higher immunity to noise and spoof attacks and smaller FAR/FRR figures. The advantages of fusion techniques make this approach attractive in biometrics, in spite of presenting higher storage requirements, processing time and computational complexity.

1.4.3.1   Feature-Level Fusion

Feature-Level Fusion may combine data from different sensors, so as to utilize the benefits of each type of sensors and mitigate the sensor's reliability issue. This method is usually known as multi-modal and can be seen in [70] where features from gait and foot pressure are concatenated. Also, face and gait features were fused by Chellapa et al. [71], Hossain et al. [64] and Liu and Sarkar [20].

Alternatively, the same data source can be used to produce different features using different processing techniques as in [72] for the fusion of static and dynamic gait features. This usually results into a higher dimension data, which can be, then, fed into a dimensionality reduction module.

1.4.3.2   Score-Level Fusion

The score is an indication of how close each feature is from the training level. Scores usually need to be normalized before fusion is possible. Scores may be then combined using the sum, max, min or product rules. Zhang et al. [39] have used fusion at score-level of features extracted using Procrustes Shape Analysis algorithm. Scores were normalized using *tanh* technique.

1.4.3.3   Decision-Level Fusion

In decision fusion, the final decision is based on the combination of decisions from multiple classifiers, or recognition modules. One advantage of decision fusion is that it picks the correct decisions made by single classifiers and combines them producing a more accurate over-all decision. It provides a robust performance against challenges that each classifier has to deal with. Decision-level fusion becomes more attractive and useful when training samples are insufficient. There are different techniques for decision-level fusion such as; majority voting [73], Bayesian Decision Theory [74], Neural Networks [75] and the Dempster-Shafer theory of evidence [76, 77]. Majority voting, however, seems to be simpler and easier to implement.

1.4.4 Tensor-based Data Representation

As seen in section (1.3.4), Multilinear Subspace Learning (MSL) (or Multilinear Projection) is the method of choice for dimensionality reduction of high-order tensorial data. In this section, we shall go through a brief review of the major contributions in this field.

Generally, MSL algorithms fall into one of two classes:

1. Unsupervised Multilinear Subspace Learning, which does not require labels for the training samples. Major methods in this class are multilinear extensions of the classical linear PCA algorithm.

2. Supervised Multilinear Subspace Learning, which uses the classes of the training samples. Major methods in this class are multilinear extensions of the classical linear LDA algorithm.

1.4.4.1 Unsupervised Multilinear Subspace Learning Methods

1. Unsupervised MSL through TTP

A two-dimensional PCA (2DPCA) [78] has treated images as matrices rather than vectors. Dimensionality reduction is achieved by projecting images to lower dimensional matrices using 2-mode TTP (N=2). Generalized Low-Rank Approximation of Matrices (GLRAM) [79] is a more generalized 2nd-order MSL method, compared to 2DPCA. It performs two linear transformations (1-mode and 2-mode), projecting the input image into a low-dimensional matrix. Later the Concurrent Subspace Analysis (CSA) [80] was applied on GLRAM to generalize projections for higher-order tensors (N>2). The Generalized PCA (GPCA) proposed in [81] is similar to the GLRAM approach, except in that it centers the data before projection (similar to PCA), which makes it more suitable for recognition applications. However, it only works on matrices.

A more generalized variation of GPCA is the Multilinear PCA (MPCA) proposed in [4]. It also addresses the issues of initialization, convergence and the determination of the subspace dimensionality. Since then, some variations of the MPCA algorithm have been reported in literature. Baysiean Tensor Analysis (BTA) in [82], which introduced graphical models of tensors, and is considered a probabilistic extension of MPCA. Robust MPCA (RMPCA) in [83] which performs the iterative solution based on

Lagrange multipliers. Non-negative MPCA (NMPCA) [84] preserves the non-negativity of the original tensors.

Another method, Incremental Tensor Analysis (ITA) in [85], is considered an incremental version of CSA. It works on summarizing high-order tensors, and updates them incrementally.

2. Unsupervised MSL through TVP

There is no much work under this category. One method was introduced in [86] as the Tensor Rank-One Decomposition (TROD). This algorithm projects images (2nd-order tensors) into low-dimensional vectors by performing least-square error, without centering of data. Uncorrelated MPCA (UMPCA) in [87] is a multilinear extension of PCA in that it extracts uncorrelated features by maximizing variance within tensors by means of successive iterative steps of TVP.

1.4.4.2 Supervised Multilinear Subspace Learning Methods

1. Supervised MSL through TTP

Two-dimensional LDA (2DLDA) was introduced in [88] to project images into low-dimensional matrices, yet using matrix-based discrimination criterion as the scatter ratio. Later, a higher-order extension of 2DLDA was introduced in [89] as the Disriminant Analysis with Tensor Representation (DATER), which uses tensor-based scatter ratio. Both 2DLDA and DATER do not converge over iterations. The General Tensor Discriminant Analysis (GTDA) algorithm was proposed in [90] as a variation of DATER that maximizes the tensor-based scatter difference by adding some tuning parameters. Compared to 2DLDA and DATER, GTDA converges well.

2. Supervised MSL through TVP

The Tensor Rank-One Discriminant Analysis (TR1DA), appears in [91], was derived from the TROD algorithm, yet with general application to higher-order tensors. Thus, it projects tensors into low-dimensional vectors, by maximizing the scatter difference.

Similar to the UPMCA algorithm in unsupervised MSL, an Uncorrelated Multilinear Discriminant Analysis (UMLDA) introduced in [92]

aims to extract uncorrelated discriminative features from tensors by using TVP to maximize the scatter ratio criterion. The solution is, thus, of sequential iterative nature. The method also adds aggregation scheme for better generalization.

Table 3 summarizes the major Multilinear Subspace Learning (MSL) methods as introduced earlier:

Table 3: Comparison of major MSL methods

Unsupervised MSL					
TTP				TVP	
LSE Minimization		Variation Maximization		LSE Minimization	Variation Maximization
$N=2$	$N>2$	$N=2$	$N>2$	$N=2$	
GLRAM	CSA	2DPCA GPCA	MPCA	TROD	UMPCA
	ITA		BTA RMPCA NMPCA		
Supervised MSL					
TTP			TVP		
Scatter Ratio Maximization		Scatter Difference Maximization	Scatter Ratio Maximization	Scatter Difference Maximization	
$N=2$	$N>2$				
2DLDA	DATER	GTDA	UMLDA	TR1DA	

1.5 Research Methodology

Human gait recognition can be typically implemented using the system of Figure 10 which includes:

a. Data Acquisition: In our case, it is the collection of video sequences of walking humans in different conditions, captured by video camera(s).

b. Pre-processing: Using image processing techniques to enhance the frame images or highlight the subject of interest. The techniques are like segmentation and filtering. In our case, we used the Accumulated Prediction, Accumulated Flow and Gait Energy to process and prepare for extraction the gait video sequences.

c. Feature Extraction: Typical methods to construct the feature vectors/tensors of the pre-processed data. In our work, we used image projection and MPCA algorithms.

d. Training (Mapping): which creates a mapping model of the extracted data and label them.

e. Classification: The process which assigns a pattern (class) corresponding to every feature vector (category) using the learned models and the features. In our model, we used the simple Linear Discriminant Functions (LDF) and K-Nearest Neighbor (KNN) classifiers
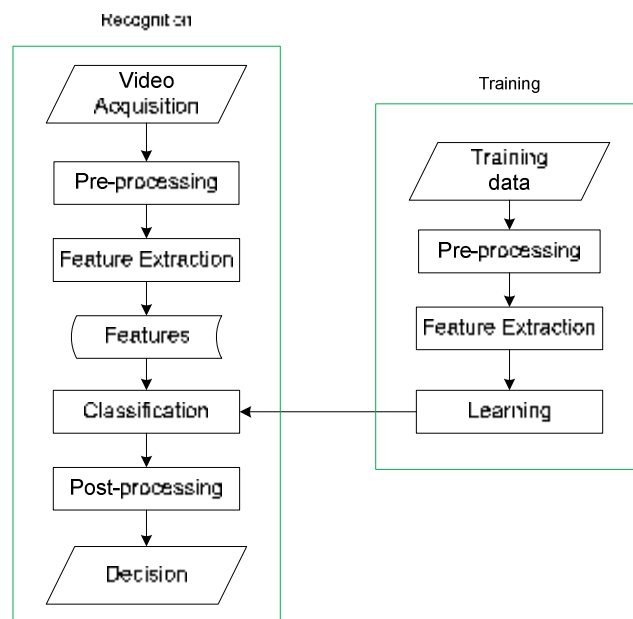
Figure 10: A typical block diagram of Gait Recognition System

Since our major objective of this study is to design a robust and covariates-independent gait recognition system, we need to carry-out different tasks and experiments. These tasks are generally divided between examining previous work, designing and evaluating new algorithms. The following are the main tasks planned and executed towards the fulfillment of the thesis:

a. Conducting a thorough study of the previous work done on the topic, and examining the different approaches suggested and/or tested.

b. Testing three different gait databases, including CASIA A, CASIA B and the locally-collected AUS database. Finally, we decided to use CASIA B dataset for the rest of the work. This dataset studies the covariates of view point,

clothing and carrying condition (as described in section 2.1.1). All video sequences were converted to still images and arranged in folders before using them in the succeeding experiments.

c. Carrying out numerous experiments on the Accumulated Prediction Image (API) method for optimization purpose. Different feature extraction and classification techniques were used in the evaluation experiments.

d. Developing new methods, later known as Accumulated Flow Image (AFI) and Edge-Masked Active Energy Image (EMAEI). These were used along with API as feature sources in our final model.

e. Testing and evaluating different feature extraction methods. Later, we have picked the MPCA, and its MPCALDA variation, as effective tensor-based methods, in addition to the linear image projection technique.

f. We didn't set a target to develop new classification methods. We only used the available tools.

g. The option of using fusion techniques in order to improve the overall system's performance and accuracy. Later, decision-level fusion schemes were chosen for the final setup. Unweighted and weighted schemes were tested and compared.

1.6 Thesis Outline

The rest of this report is organized as follows: Chapter 2 introduces the main pre-processing techniques used for gait representation including API, AFI and EMAEI. Chapter 3 explains the methods used for feature extraction and dimensionality reduction including image projection and MPCA algorithm. Chapter 4 demonstrates the base classifiers as well as fusion techniques implemented for the final decisions. And we present our experimental results with analysis in Chapter 5. The thesis ends with conclusions and recommended future work in Chapter 6.

Chapter 2: Gait Preprocessing Techniques

In this chapter, we explain in more details the database used in our research, CASIA B; its collection, structure, size and arrangement of its data. We also look at some of the major gait preprocessing techniques used in literature, and the modifications and/or optimizations implemented to those as part of our analysis.

Initially, our analysis focused on the Accumulated Prediction Image (API) representation, and this was mainly due to the promising results achieved in [3]. The representation performed well in producing representative and distinctive features. The method is explained in more details in section 2.2. Tens of different experiments have targeted to enhance the representation power of the method, and to test or develop different feature extraction and classifiers aiming for the highest recognition rates possible. As research work progressed, other state-of-the-art gait representations were tested. And aided with an extensive literature review, we conclude that spatio-temporal (accumulated error) and energy-based methods are more robust and have been used to report most of the highest average recognition rates, as in [13-16, 34], compared to other methods. One main advantage of using silhouettes is that it offers huge amount of data in a single template, which, on the other hand, can be also regarded as additional computational cost, a matter that necessitates preprocessing of such data. Hence, our subsequent analysis has solely focused on accumulated error and energy-based methods, mainly Gait Energy Image (GEI), Accumulated Flow Image (AFI) and Active Energy Image (AEI). This is of course in addition to the API gait representation method. As detailed later in this chapter, we will look at the differences between GEI and AEI. After quick experiments on both methods, we have dropped GEI from our study. Although improvements may be expected, and in spite of the fact that GEI, or its modified versions, has become one of the most accepted and used energy-based gait representation in the literature, as in [35-38, 93-95], we had the intuition that AEI may be more promising in our analysis as it better captures dynamic features. As for API and AFI, our plan was to exhaust each of the two methods as to the best achievable individual recognition rates, by varying their different parameters, and testing them against different feature extraction techniques. Optimization of API and AFI is further discussed in sections 2.2 and 2.3 respectively.

2.1   Gait Sequences and Cycles

2.1.1 CASIA B Dataset

We have picked the CASIA database, particularly dataset B, for the sake of our experiments and analysis. This dataset consists of 124 subjects, each subject walked 10 times resulting in 10 video sequences; 6 sequences in normal condition, 2 sequences carrying different shapes and sizes of bags, and 2 sequences wearing different types of coats. With 11 cameras distributed at 11 different view angles, ranging between 18° and 180°, we end up with a total of 13,640 video sequences. As we only focused on testing for the carrying condition and clothing, we have used sequences corresponding to only one view angle, which is 90° or lateral view. This means we have actually tested 1240 sequences. The main reasons for choosing this database are therefore:

a. It is still currently the largest gait database collected and used in the literature, which means it can be split into different groups for different verification tests, seeking better generalization of the designed method.

b.   It comprises subjects of both genders and different ages, which again makes the experiments closer to realistic scenarios.

c. It tests not only for the change in carrying condition and clothing, but also for the change in the view angle. Although our methodology is designed to fix the view angle, we expect that future works should extend to test all other view angles as a robustness test.

d. It was widely used to report the highest recognition rates with the state-of-the-art gait recognition methods, as in [13-16, 34], which would, then, give us the chance to fairly compare our method's performance.

2.1.2 Database breakdown

To start with, we had to convert all video sequences to still images (or frames) and arrange them in folders. Each video sequence was converted to bmp images in a 25fps conversion scheme. We dropped the blank frames and we only kept those frames in which the full subject is visible, resulting in an average of 60 frames per sequence, and yielding at least a total of 75,000 images. These images were, then, arranged in folders; 1 main folder per condition, or '*normal*', '*bag*' and '*coat*'. Inside

each of these folder images were distributed in sub-folders corresponding to the subjects' numbers, and then the sequences' numbers.

2.1.3 Gait Cycles

As seen in Figure 1 (section 1.3.2), one gait cycle is measured between two successive heal strikes. For example, the walking duration from the right heal strike to the next right heal strike. But this can generally be any full repeated walking pattern. One gait sequence should span over one or more gait cycles. Usually, the more cycles we have in a sequence the more patterns can be extracted. Most of the reported gait recognition methods have considered the gait cycle, or half-cycle, as one gait sample. Therefore, they developed algorithms to detect those cycles. One of the most widely adopted methods in gait cycle detection (or gait period detection) is developed in [11]. It is based on the summation of the foreground pixels in each silhouette (or the bottom half of the silhouette) in the sequence. This number is minimum when legs overlap and maximum when they are at full stride. Followed with distance computations, gait cycles can be estimated.

We have developed a new algorithm for gait cycle detection based on a different criterion, which is Convex Image (Figure 11). Convex Image (or value) for each gait image was generated. The silhouette corresponding to the full stride produces the largest convexity value. Similarly, when the two legs overlap, this produces the smallest convexity value. Our method is simpler compared to the one in [11]. After an experimental comparison between the use of the number of foreground pixels and the convex value, our method proves to be more accurate in detecting cycles. In other words, the convex image property is more effective than the measurement of the distance between the two legs or the total number of pixels within that same area.

However, and after using this technique in our initial experiments, it scored less compared to the case when we input the whole gait sequence as one sample. This means we have actually omitted this module from the recognition system as deemed unnecessary. In fact, we thought later that feeding the whole sequence as one sample may help in extracting more features intrinsic in the multi-cycle's image. Besides, the proposed gait cycle detection methods work effectively only for the lateral-view

silhouettes, and hence cannot be generalized to test against different view angles in the future.
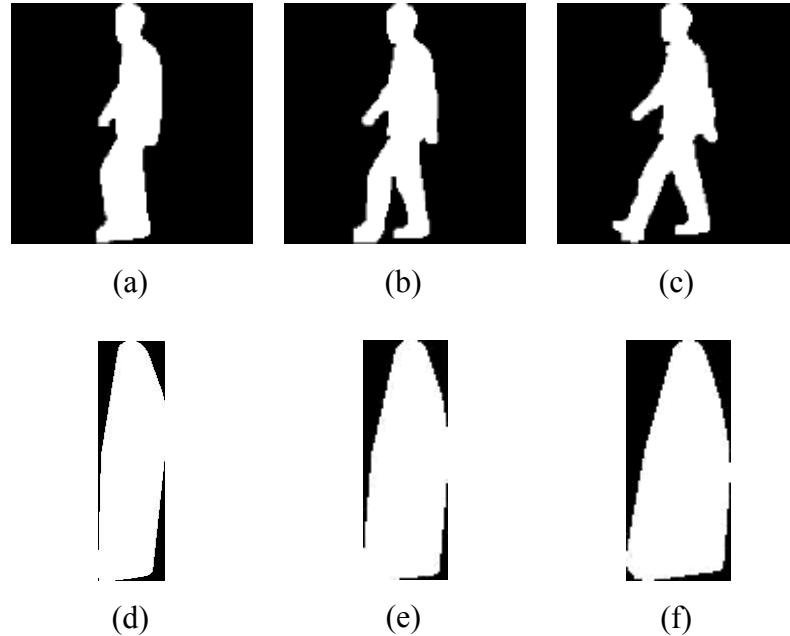
Figure 11: (a), (b) and (c) The first three silhouettes for subject 1, normal sequence 1, and their corresponding Convex Images, (d), (e) and (f) respectively.

2.1.4 Design Baselines

a. In our experiments, we have considered all the subjects available in the database, i.e. 124 subjects. This is contrary to some papers in literature where they actually dropped a few subjects for having incomplete or low-resolution silhouettes. In [15, 96], for example, only 120 subjects are used. In [66], only 98 subjects were used. One of our goals was also to test the system's robustness against such flaw factors.

b.   Resolution: The resolution of the extracted frames is 240x320 pixels. The raw silhouettes available in the database were of resolution 240x240 pixels. In API and AFI methods, we work on the original frames. These were cropped to 168x320 to remove the areas of no motion and, thus, reduce the noise generated in the background. These resolutions were used in the recognizer implementing image projection and LDF. When implementing the multidimensional (tensor-based) algorithms, the resolutions were reduced to 80x42 pixels. As for the EMAEI method, we have used the original resolution

of 240x240 in the first recognizer (Projection + LDF), and reduced it to 64x64 in the MPCA and MPCALDA methods. It is noticed that the image resolution has insignificant impact to the final recognition rate. Y. Dupuis et al. have also shown in [34] that the final recognition rate is almost independent of the silhouettes' resolution. We have picked 64x64 resolutions as it produced recognition rates close to those resulted from using the original resolution (240x240), while still clearly better than those resulted from using the resolution 32x32 or lower. Therefore, this selection of resolution should have negligible impact to the end result. Besides, it gives us the chance to test the system against lower-resolution images, which is actually one main advantage of gait recognition system.

c. As said in section (2.1.3), we have dropped the gait cycle detection stage from our system, which again should improve the computational power. Additionally, we noticed that breaking the sequence into cycles does not help much in our selected gait representation methods, as they are all accumulating methods and are expected to accumulate all frames, extract the representative features and detect the patterns in the overall accumulated image, where cycles are actually intrinsically represented.

2.2   Accumulated Prediction Image (API)

2.2.1 Description of the API method

The construction of this image (or images) is derived from the prediction error computation technique used in digital video coding. Each frame is subtracted from its immediate previous (past) frame, a process known as forward prediction. This subtraction results in a Prediction Error Image. A threshold is, then, applied to the image so as to filter out those pixels that are non-motion related. Next section details on the process of optimization of the threshold value selection. The resulted prediction error images after thresholding are, then, accumulated resulting in the Accumulated Prediction Image (API). For further filtration of the non-motion (or noise) pixels, a decision should be made whether to use forward prediction or backward prediction (difference between each frame and the immediate future frame). This decision is based on the minimization of the Sum of Absolute Difference (SAD)

between the forward prediction error image and the backward prediction error image. The image that minimizes this SAD value is taken towards the final construction of the API. Besides, and for the sake of better representation, we have split the process into two parts, positive API (using even-numbered frames) and negative API (using odd-numbered frames). As such, each video sequence ends up with two APIs (Figure 12) that can be used for feature extraction. Note that we have maintained the spatial information in the two images, a matter that shall be of added value in feature extraction.

(a)   (b)
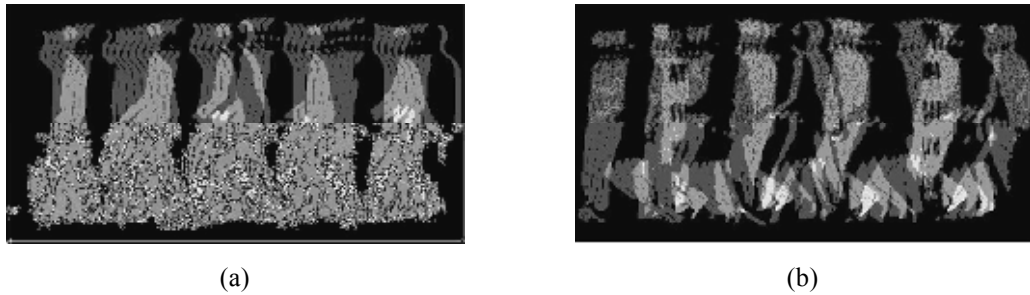
Figure 12: (a) Positive API, (b) Negative API

API method was successfully implemented in [3] to report as high recognition rates as 100% on a locally-collected gait database. It was noticed that this gait representation, followed with the feature extraction technique explained in section (3.1), have contributed in extracting highly linearly-separable features that were, then, fed into a linear classifier, in a relatively simple approach and of low computational cost.

2.2.2 Optimization of the API representation

As seen in 2.2.1, this method is dependent on thresholding. Therefore, it was intuitive to think of optimizing the threshold value that is responsible for the quality of the output gait image. The tested and optimized parameters are:

a. Thresholding factor: In principle, this can be any digit between 1 and 255. Obviously, however, we want to avoid high values that would result in waiving some representative features. At the same time, we wish to remove the intrinsic noise (or non-motion values). At initial run of the codes we have noticed that 10 should be the maximum value to test for. Beyond this, we start losing big portion of the image, and thus, producing considerably lower

recognition rates. This can be understood if we know that the intensity values from 1 to 10 represent almost 98% of the image. Therefore, the parameter was varied between 1 and 10, with a step value of 1, and tested accordingly. The final optimum value was 7. Figure 13 below shows sample prediction images resulted from the first two frames in the first normal walking sequence, and using threshold values of 1 and 7.

(a)   (b)

(c)   (d)
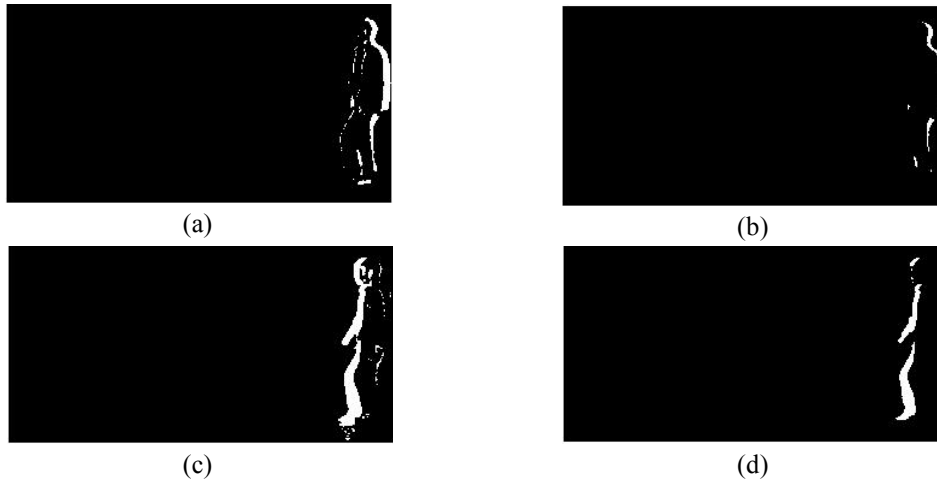
Figure 13: (a) Positive prediction with threshold factor of 1, (b) Positive prediction with threshold factor of 7, (c) Negative prediction with threshold factor of 1, (d) Negative prediction with threshold factor of 7

b. The averaging factor: This parameter is merely an additional control (fine tuning) factor that may contribute in a more optimized filtration. The value was varied between 0.1 and 1.0 with a step value of 0.1. Eventually, 100% of the average value (that is value of 1) is considered in our calculations, as proving to produce the best results.

2.3   Accumulated Flow Image (AFI)

2.3.1 Optical Flow

Optical flow corresponds to the movement of pixels in a sequence of images [97], or in other words the rate of change of motion (velocity) of intensity patterns. Such information is useful in the gait problem. It can be used in objects' segmentation and motion detection and would produce spatial arrangement that should provide a pool of features for object classification. In literature, there are different methods for

the computation of optical flow, including methods of differentiation, correlation, feature-based methods and hierarchical approaches. Differentiation methods are, although classical, still the popular ones. These include global differentiation techniques such as Horn and Schunk, and local differentiation techniques such as Lucas and Kanade.

In AFI representation, images are constructed using the Horn and Schunk differential method [98]. Therefore, we would introduce herein its main equations, constraints and assumptions, while the derivation remains out of the scope of this study.

First assumption made in the calculation of optical flow is that the image brightness, reflectivity and illumination are constant while moving in a short time interval Δt, from $t1$ to $t2$. Let $I(x,y,t)$ be the image intensity at point (x,y) and time t. Since I is constant at a fixed point, that is

$$\frac{dI}{dt} = 0 \tag{1}$$

By chain rule, we get:

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = 0 \tag{2}$$

Let:

$$u = \frac{dx}{dt} \ and \ v = \frac{dy}{dt}$$

Where u and v are the two components of the flow vector $\mathbf{v}$, which we are after in the computation of the optical flow. Equation (2) can be rewritten, then, as:

$$I_x u + I_y v + I_t = 0 \tag{3}$$

Equation (3) becomes the first constraint in Horn and Schunk method. The second constraint is the smoothness constraint. The idea is that the optical flow should vary smoothly everywhere over the entire image. This can be derived by minimizing the square of the magnitude of the optical flow vector $\mathbf{v}$:

$$E_c^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \tag{4}$$

This equation represents the difference between the flow vector and its neighbors. Adding this to equation (3), or the constraint equation, enables us to solve for u and v.

After a method of derivation explained in [98], we get two equations to iteratively update for the values of u and v:

$$u^{n+1} = \bar{u}^n - \frac{I_x \left[I_x \bar{u}^n + I_y \bar{v}^n + I_t \right]}{\alpha^2 + I_x^2 + I_y^2} \tag{5}$$

$$v^{n+1} = \bar{v}^n - \frac{I_y \left[I_x \bar{u}^n + I_y \bar{v}^n + I_t \right]}{\alpha^2 + I_x^2 + I_y^2} \tag{6}$$

where; u^{n+1} and v^{n+1} are the new optical flow components at iteration $n+1$, $\bar{u}^n$ and $\bar{v}^n$ are the averages of the optical flow vector components in a small neighborhood, and at previous iteration n. I_x and I_y are the spatial gradients. I_t is the temporal gradient, and α^2 is a weighing factor

2.3.2 Description of the AFI method

The construction of the AFIs is based around equations (5) and (6) being an iterative method. As such, the parameters that govern its behavior are: number of iterations, the smoothness factor α and initial values of u and v. In our experiments we assume initial values of 0 for each of u and v components. This leaves us with two parameters to control only. More details on this selection process are found in next section.

The computation of the optical flow between every two consecutive frames produces a Flow Error Image. Similar to our work in API (section 2.2.1) we accumulate these error images towards the construction of the final AFI. In this case, however, we have to deal with two sets of variables every time, u and v. Recalling that we actually split the process in two parts, positive image and negative image, means we'll have 4 images after every run of the experiment. After the initial experiments, we found that the horizontal component (u) produces higher quality images, yielding higher recognition rates. Therefore, in all our subsequent analysis we have only considered the u-based AFIs. Every mention of AFI herein, in this paper, would implicitly assume using the u component only. Figure 14 shows sample AFIs constructed from the first sequence of subject 1 in the normal conditions.

<div align="center">(a)   (b)</div>
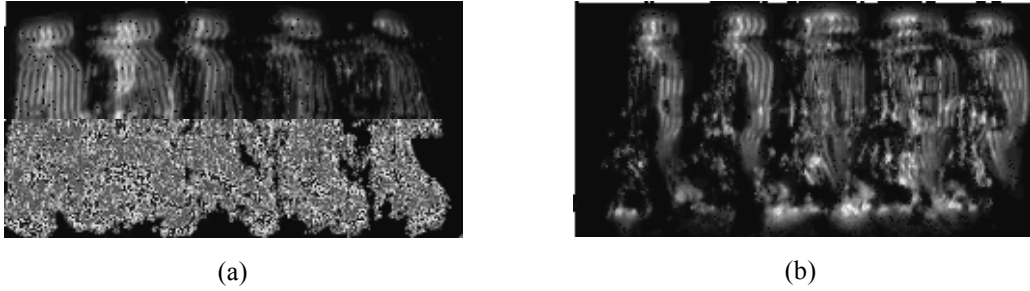
Figure 14: (a) Positive AFI, (b) Negative AFI

Our work on optical flow was inspired by an earlier adoption of this technique by Lam et al. [63], and then by Wang et al. [96]. Their method, called Gait Flow Image (GFI), although uses the same principle of optical flow, is different in the following attributes:

- GFI is based on silhouettes, unlike our method, AFI, that operates directly on the original images, and then automatically thresholding. By working on the images we anticipate to better preserve their properties.

- GFI required the computation of the gait periods (cycles), whereas AFI accumulates all the frames in only two images and as we shall see later in feature extraction, features from these two images are interleaved, leaving only one sample vector per subject. In other words, every full sequence corresponds to only one sample. This is expected to increase the computational efficiency, without compromising features quality, as cycles are still embedded in the AFIs.

- In GFI, Silhouettes are normalized to fixed size before being aligned centrally around their horizontal centroids. In AFI, we maintain the original spatial information, producing strip of images that should be a good interpretation of the subject's motion.

2.3.3 Optimization of the AFI representation

Referring to the previous section, we know that the number of iterations and the smoothness factor are the most important parameters in defining the quality of the output image. As such, these two parameters have been varied and method tested accordingly. Initial values of the optical flow vectors, u and v, are kept at default 0.

a. Number of iterations: These were varied between 1 and 100 but only at randomly selected values; 1, 5, 10, 20, 32, 64 and 100. Results were best at 64, and so it was fixed for all succeeding experiments.

b. Smoothness factor: This value is responsible for the quality of the output gait image and how representative it is for motion patterns. The higher the value, the smoother the image. However, at some relatively high values, smoothness leads to fewer representative features. At initial random selection of values, it was noticed that values above 20 should be avoided. The parameter was, thus, varied between 1 and 20 with a step value of 5. The value corresponding to the optimum result was 5, and so it was fixed for the succeeding experiments.

2.4 Gait Energy Image (GEI)

2.4.1 Description of the GEI method

This method does not associate to our final model. However we explain it here briefly for the following reasons:

a. It was one of the earliest methods to introduce the concept of accumulating silhouettes in the production of energy images that represent the body's statics and dynamics. In other words, an appearance-based method.

b. Its significance and being, or variations of the original method, widely adopted in various gait research papers.

c. To compare it with another method, namely Active Energy Image (AEI) that represents the basis to our next method, Edge-Masked Active Energy Image (EMAEI).

The method, introduced in [93], simply accumulates the summation between each consecutive preprocessed, normalized to a fixed size and aligned around horizontal centroid, silhouettes and then average them by the total number of frames in the cycle, or:

$$G(x, y) = \frac{1}{N} \sum_{t=1}^{N} B_t(x, y) \tag{7}$$

Where, $B_t(x,y)$ is the preprocessed silhouette image at time t, and N is the number of frames in a complete cycle of the gait sequence.

This process produces one single image for the full gait cycle contributing to a lower computational complexity and storage space. Sample GEI images, of the first subject in the three datasets, are shown in Figure 15.

2.5 Masked Active Energy Image (MAEI)

2.5.1 Active Energy Image (AEI)

Recall that the GEI method produces a single gait sample image per cycle, in which the intensity at a single point is a representation of the frequency of that particular part of the body. We notice however that GEI accumulates both dynamic and static features of the moving body, which makes it appearance-based and sensitive to the change in clothing and carrying conditions. Zhang et al. have introduced in [13] the AEI gait representation that better highlights the dynamic features of the subject and eliminates those related to the static parts, especially the bag and the coat in our case.

This was achieved by simply accumulating the differences (instead of summation as in GEI) between every two consecutive frames and averaging them by the number of frames N, or:

$$A(x, y) = \frac{1}{N} \sum_{t=0}^{N-1} D_t(x, y) \tag{8}$$

Where $D_t(x,y)$ is the difference between frames given as:

$$D_t(x, y) = \begin{cases} f_{t(x,y)}, & t = 0 \\ \|f_t(x, y) - f_{t-1}(x, y)\|, & t > 0 \end{cases} \tag{9}$$

Where $f_t(x,y)$ is the tth silhouette.

Example AEI images are shown in Figure 15 and compared to GEI for the same subject (Subject 1) taken from the same dataset.

2.5.2 Description of the MAEI method

An observation of the AEI images, for CASIA B dataset, could tell us that the mid portion of the images seem to have the least dynamic features. AEI's major contribution was actually the minimization of the static traits linked with the bags and coats in the sequences. It was intuitive to think about tracing those areas of less

contribution to the final features and masking them out, and expecting this to improve the final recognition rate when testing for the bag and coat sequences. Therefore, we have opted to test a naive model that would impose the mask manually during the formation of the AEI yielding, eventually, a masked AEI or MAEI. Obviously this method is based on observation of particular datasets. However, and for the sake of our study, this was sufficient. If the intuition is right, a more efficient system can be developed as part of the future work, so as it would automatically trace the least significant features and mask them out, being independent of the datasets. This can be accomplished using filtering, wrapping or decision tree approaches. A good example of feature subset selection techniques was presented in [34].

We have tested two strategies of masking:

a. *Zero Masking*: which simply replaces-by-zero all values that lie between two predetermined lines that sandwich the least significant features (or least intense pixels). The two lines were chosen at rows 30 and 155. Knowing that all AEI images are of fixed normalized size 240x180 pixels, means this fixed selection should generalize well as a primary mask. Example MAEI images using Zero Masking are shown in Figure 15, images g, h and i.

b. *Edge Masking*: Instead of cancelling all pixels between our reference lines, we have applied the edge detection technique. Our intention was to further minimize the contribution of the static features related to the bags and coats, while still maintain some discrimination power in them. Edge detection, using 'Sobel' edge operator, is applied to every difference image, before images are accumulated. This method performed better compared to the Zero Masking. We notice from Figure 15 that although the Edge-Masked AEIs (or EMAEIs), j, k and l, show little information in the mid portion, are still non-zero, indicating those areas with motion difference.
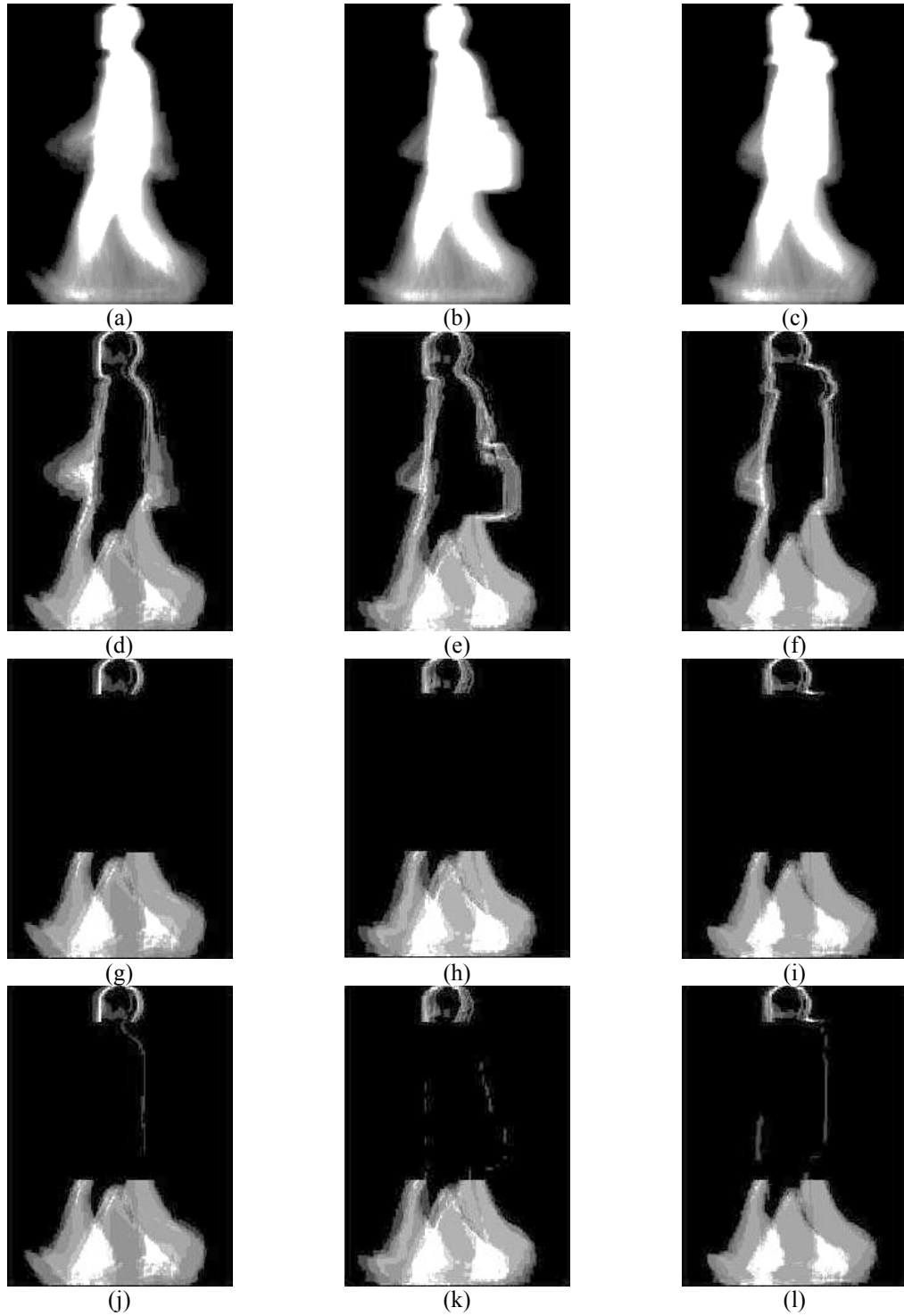
Figure 15: (a) GEI of subject 1 under normal condition, (b) GEI of subject 1 with bag, (c) GEI of subject 1 with coat, (d) AEI of subject 1 under normal condition, (e) AEI of subject 1 with bag, (f) AEI of subject 1 with coat, images (g, h, i) are the corresponding MAEIs to images (d, e, f) using Zero Masking, images (j, k, l) are the corresponding MAEIs to images (d,e,f) using Edge Masking.

Chapter 3: Feature Extraction and Dimensionality Reduction

3.1   Image Projection + 1D DCT

It is a simple approach and we used it in combination with a linear classifier as we shall see later in Chapter 4. It takes the horizontal projection of each gait image, followed by one-dimensional Discrete Cosine Transform (DCT) to smooth and reduce the size of the projected values. This produces one feature vector which is truncated using a pre-selected cutoff value which is determined empirically. After testing the code through a range of cutoff values between 20 and 200, the value 100 was, then, the selected cutoff value used in all succeeding experiments. When applying this approach to the APIs and AFIs, being two images per sample, we use ½ the cutoff value to truncate the vector for each image. Then the two feature sub-vectors are interleaved to produce one combined vector of dimensionality 100. This approach is summarized in Figure 16. When used with the EMAEI gait image, the whole image produces one single feature vector of dimensionality 100.
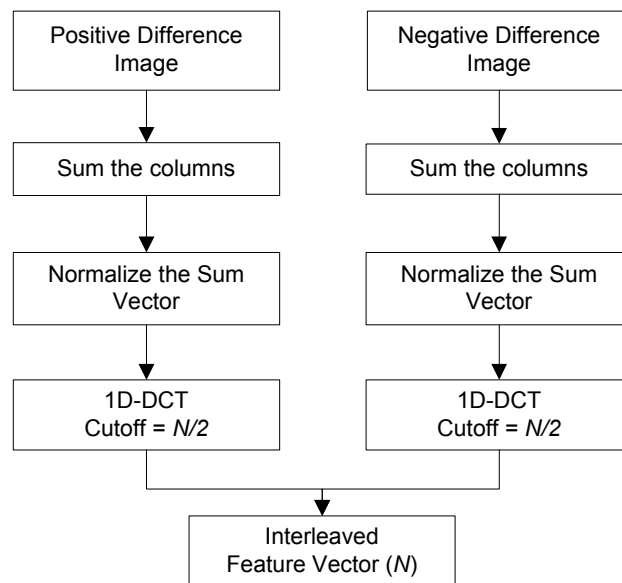
Figure 16: Image horizontal projection method for feature extraction. The flowchart shown is for the two cases of API and AFI. EMAEI requires only one branch, for a single image.

3.2   Multi-linear Principal Component Analysis (MPCA)

As explained in sections 1.3.4 and 1.4.4, MPCA is an unsupervised Multilinear Subspace Learning (MSL) method that implements the Tensor-to-Tensor Projection (TTP) technique, to project high-order tensor objects into lower-dimensional tensors. In other words, it applies dimensionality reduction directly on objects in their tensorial form. The main advantages and reasons to choose this method, as introduced in [4], amongst other MSL methods are:

1. It is generalized to work on higher-order tensors. Thus, it introduces systematic procedure and provides tensor representation to tensorial objects, compared to other heuristic methods or those that use vector or matrix forms.

2. It deals with some issues related to the iterative solution, like initialization, termination, convergence and the determination of the projection dimensionality. It was successfully used in [4] to produce better recognition results without optimization.

3. It was implemented on gait recognition, tested on the USF database and compared to the HumanID algorithm introduced in [11]. It was, thus, a good benchmark to develop our method and compare our results.

4. It is noteworthy here that we have also tested another MSL method that was proposed after the MPCA, and by the same authors, in [92]; the Uncorrelated Multilinear Discriminant Analysis (UMLDA). Although the method was used to report better average recognition rates compared to the MPCA approach, it is more complicated in that;

   a. It uses the Tensor-to-Vector Projection (TVP) technique that incorporates sequential EMP operations.

   b. It is more sensitive to the Small Sample Size scenario. And to overcome this problem, a regularization approach was added to the main algorithm.

   c. It incorporates an additional aggregation module to overcome the issue of un-optimized iterative solution, compared to the MPCA that performs relatively well without optimization.

   d. When tested specifically on gait sequences, it outperforms the MPCA only for some of the probes and cannot be expected to generalize well.

3.2.1 Description of the MPCA algorithm

The objective of the MPCA algorithm is to define N multilinear projection matrices $\{U^n \in R^{I_n \times P_n}, n = 1, \dots, N\}$, each of which is to map an input tensor object $X_m \in R^{I_1 \times I_2 \times \dots \times I_N}$ into a lower-dimensional tensor object $Y_m \in R^{P_1 \times P_2 \times \dots \times P_N}$ (where $P_n < I_n$ for n=1, 2, …, N).

The MPCA objective function is defined as:

$$\{U^n, n = 1, \dots, N\} = argmax_{U^1, U^2, \dots, U^N}(\psi_y) \qquad (10)$$

Where ψ_y is the total tensor scatter, defined as:

$$\psi_y = \sum_{m}^{M} \|Y_m - \bar{Y}\|_F^2 \qquad (11)$$

Where M is the number of samples, and $\bar{Y}$ is the mean tensor calculated as:

$$\bar{Y} = \frac{1}{m} \sum_{m}^{M} Y_m \qquad (12)$$

Equation 10 means that the objective of MPCA is to determine the N projections that maximize the total tensor scatter ψ_y

Below is a general pseudo-code of the MPCA algorithm:

Input: $X_m \in R^{I_1 \times I_2 \times \dots \times I_N}, m = 1, \dots, M$

Algorithm:

1. Center the input samples, $\tilde{X}_m = X_m - \bar{X}$, where $\bar{X} = \frac{1}{m}\sum_{m=1}^{M} X_m$ is the sample mean.
2. Calculate the eigen decomposition of $\phi^n = \sum_{m=1}^{M} \tilde{X}_{m(n)} \cdot \tilde{X}_{m(n)}^T$
3. Set $\tilde{U}^n$ to consist of the eigenvectors corresponding to the most significant P_n eigenvalues.
4. Calculate cumulative distribution of eigenvalues, $\lambda_{i_n} = \lambda_{i_{n-1}} + \lambda_{i_n}$
5. Determine the dimension of the projected space P_n
6. For $k=1:K$ (K: number of iterations)
   - For $n=1:N$

     Set $\tilde{U}^n$ to consist of the P_n eigenvectors of ϕ^n corresponding to the largest P_n eigenvalues.

7. Calculate the weight tensor (to be used in classification):

$$\mathcal{W}(p_1, p_2, \dots, p_N) = \sqrt{\prod_{n=1}^{N} \lambda_{p_n}^{(n)}}$$

8. Calculate the projection of tensor samples:

$$\mathcal{Y}_m = \mathcal{X}_m \times_1 \tilde{U}^{(1)T} \times_2 \tilde{U}^{(2)T} \dots \times_N \tilde{U}^{(N)T}, m = 1, \dots, M$$

Output: $\mathcal{Y}_m \in R^{P_1 \times P_2 \times \dots \times P_N}, m = 1, \dots, M$

Finally, the output tensor is vectorised and sorted by discriminality. The number of vector components kept for the analysis is determined empirically. We have tested values ranging between 200 and 1000, with a step value of 100. The value that yielded the best average classification rates using different gait inputs was 600.

3.2.2 Tackling the iterative solution issues

In an attempt to optimize the iterative solution, the following techniques were adopted in [4]:

1. Initialization by Full Projection Truncation (FPT): This is done by starting iterations in each mode n by assuming $P_n = I_n$. In each mode, we get I_n number of eigenvalues. These are sorted in descending order, each two successive eigenvalues accumulated and then normalized by dividing by the sum of all eigenvalues. The resulted accumulated and normalized eigenvalues are, then, used to determine the projection dimensionality reduction. This was used in [4] and proven to yield quick convergence.

2. Termination: is based on the objective function ψ_y so that $\psi_{y_k} - \psi_{y_{k-1}} < \eta$ should be satisfied before termination, where η is a user-defined small number threshold (e,g, 10^{-6}).

3. Determination of the tensor subspace dimensionality: using the Q-based method. Q is a user-defined number, determined empirically, that represents the number of accumulated eigenvalues (or energy) kept. If $Q=97$, for example, P_n is identified as the index corresponding to the normalized eigenvalue which is less than or equal to 0.97.

3.2.3 MPCA plus Linear Discriminant Analysis (MPCALDA)

The purpose of computing the LDA after the MPCA output is to maximize the ratio of the between-class scatter matrix to the within-class scatter matrix, in an attempt to produce higher class discriminability and, thus, higher classification rate. Therefore, we have actually tested two methods; MPCA and MPCALDA. In both sets of experiments the data setup and classifier are the same. In the case of the LDA algorithm, the dimension of the feature vector is fixed to *C-1*, where *C* is the number of classes. In our data *C*=124, yielding an LDA feature vector of dimension123.

In this case, the number of vector components kept in the analysis, and after experiments on different values, is 200.

3.2.4 Data arrangement in tensorial form

The MPCA method was applied to our pre-processed gait images explained in chapter 2 so as to extract the gait features. With three gait representation methods, namely API, AFI and EMAEI, we shall have three models for feature extraction using the MPCA algorithm and three more using the MPCALDA algorithm. For this purpose, our data need to be arranged first in tensorial form. In comparison, Haiping et al. [4] used the gait silhouette sequences as samples, more precisely each gait sample is taken as the normalized half-cycle. While in previous methods, each single silhouette represented a sample. In our approach, however, we used neither of the two representations. As we mentioned earlier in this paper, we dropped the gait cycle determination module in our work, since our accumulated gait images present these intrinsically. As such we chose the gait samples to be the accumulated images mentioned above. With these arranged in tensorial form and fed to the MPCA algorithm, we introduce a novel feature representation method called Accumulated Gait Tensor (AGT). Primary comparative experiments have yielded better results using this representation compared to using single silhouettes or cycles. However, further analysis on this part, and using different gait databases and experimental set-ups, is necessary before generalizing this finding. One big advantage of the AGTs is reducing the dimensionality of the tensorial input dramatically while preserving discriminative features. Additionally, we reduce the input tensors from being three-mode (3rd-order) tensors, spatial and temporal modes, into two-mode (2nd-order) tensors since the time-mode is embedded in the accumulated gait images. Below is a

break-up of the tensorial data representation and dimensionality for each of the three gait representations for more clarifications:

3.2.4.1 API Tensorial Representation

The API approach applies to the original frames, and these are of dimension 240x320 pixels. In our preprocessing, we have cropped these to 168x320 pixels where the rest of the image has no information. Moreover, and for the sake of using them in the MPCA algorithm, we have reduced the resolution by 75% to 42x80 pixels. Since in API we have two accumulated images per subject, positive and negative, the final tensor dimension would be 42x80x2, which is still a 3rd-order tensor. However, compared to using the individual silhouettes, which average to approximately 60 per sequence, in the latter case we get a tensor dimension of 42x80x60. And using each gait half-cycle as sample, with average gait half-cycle of approximately 10 silhouettes, the tensor dimension in this case is 42x80x6. The reduction of the input tensor dimensionality is, thus, apparent. Applying the AGT concept to our data here, we shall use the term Accumulated Prediction Tensors (APTs) in our experimental analysis later to indicate the APIs in tensorial form.

3.2.4.2 AFI Tensorial Representation

AFI also uses the 240x320 frames. Therefore, the produced tensors have the same three-mode of 42x80x2. This data is called Accumulated Flow Tensors (AFTs).

3.2.4.3 EMAEI Tensorial Representation

Unlike in the two previous gait representations, EMAEI uses the gait silhouettes, and not the frames, as inputs. These are also found in CASIA B dataset with the resolution 240x320. In preprocessing, all silhouettes were cropped to produce 240x180 pixels images, in which silhouettes were centered and resized to occupy the full frame. This way, we ensure that all silhouettes are aligned. In preparation for using them in the MPCA algorithms, we have also reduced the resolution by 75% to 60x45 pixels. And knowing that we have one image to represent each gait sequence means that we have actually reduced this problem into a two-mode tensor problem, or two-dimensional (2D) PCA. For the sake of consistency with the other two tensorial representations, however, we'll call the tensor data here Edge-Masked Active Energy Tensors (EMAETs).

Chapter 4: Classification and Fusion Methods

In our study, we have invested most of the time in improving the preprocessing and gait representation techniques, and testing different feature extraction schemes; so as to produce highly discriminative features, which shall, then, make the classifiers' job easier. Experimental results (in Chapter 5) show that this target was achieved, and thus we used only the simple Linear Discriminant Functions (LDF) and K-Nearest Neighbor (KNN, K=1) methods for classification in our model. We still believe, however, that testing some more advanced classifiers, such as the Polynomial Networks, Support Vector Machines (SVM) and Neural Networks may enhance the final results. This goes as part of the proposed future work.

4.1 Linear Discriminant Functions (LDF)

We have opted to start using this naive linear classifier, for being simple, fast and robust, especially if data is linearly separable. As such, it would quickly classify data, or at least give us feedback on its behavior. In LDF, each subject represents one class, yielding N classes for N subjects. All features are grouped in a feature matrix V. A target matrix T is constructed from the Label vector, by assigning 1's for the person whose vector correspond to. Model weights can, then, be calculated by the formula:

$$W = ((V^T V)^{-1} V^T) * T \qquad (13)$$

After finding the model weights, one set of weights for every subject; we can get the score S by dot product between the weights matrix W and the testing feature matrix Y:

$$S = Y * W \qquad (14)$$

Final decision can, then, be made based on the maximum score. In other words, the class is predicted to be the matrix index corresponding to the maximum score.

4.2 1 Nearest Neighbor Classifier

This is another simple statistical classifier. The 1 Nearest Neighbor Classifier compares each data point with its closest data point in the training data and assigns it the same label. There are different measures for defining the distance between data points. Seven distance measures were tested in [4] including Euclidian, Mahalnobis, Angle and modified (or weighted) versions that can be more suitable to use with tensorial data. The distance measure we used in our classifier as have produced the highest results, is the Modified (or weighted) Angle Distance (MAD) given by:

$$d(a,b) = -\frac{\sum_{h=1}^{H} a(h).b(h)}{w(h)\sqrt{\sum_{h=1}^{H} a(h)^2 \sum_{h=1}^{H} b(h)^2}} \tag{15}$$

Where H represents tensors and $w(h)$ represents the weight tensor computed in the MPCA algorithm, and defined as:

$$\mathcal{W}(p_1, p_2, \dots, p_N) = \sqrt{\prod_{n=1}^{N} \lambda_{p_n}^{(n)}} \tag{16}$$

Where $\lambda_{p_n}^{(n)}$ represents the p_nth n-mode eigenvalue corresponding to the projection matrix $U^{(n)}$.

4.3 Voting Schemes for Decision-Level Fusion

4.3.1 Unweighted Voting (UWV)

This is the simplest and most straightforward voting method for decisions' combination. It simply counts the number of decisions for each class and assigns the sample to the class that received the highest number of votes. In this case the final decision D is given by:

$$D = arg \max_{C_j} \sum_{1}^{k} \delta(D_k, C_j) \tag{17}$$

Where C_j is the target class,

D_k is the decision of the base classifier C_k, and

$$\delta = \begin{cases} 1, if\ D_k = C_j \\ 0\ otherwise \end{cases} \tag{18}$$

4.3.2 Weighted Voting (WV)

By introducing weights to the formula, we aim to give more significance to those base classifiers that perform better individually, an approach that should improve the final decision's accuracy.

Most of the weighted voting methods are derived from evaluating the decisions of all base classifiers, sort them according to their estimated accuracy (using the training dataset), and finally give each of them a proportional weight. Accuracy is estimated by validation. For this purpose, we have split the training data used in Setup 1 (*Tr1*) into two parts; 2 *normal* sequences for training and 1 *normal* sequence to test for the classifiers' accuracy. There are different weighted voting methods. We test here one simple form of them:

After estimating the error of each base classifier e_k based on the validation data, we evaluate the authority value a_k, which is equivalent to *1-e_k*. The weight w_k is given by:

$$w_k = \frac{a_k}{\sum_i a_i} \tag{19}$$

After assigning the weights to all base classifiers, the final decision will be evaluated using similar equation to 17, by only adding the weights w_k:

$$D = arg \max_{C_j} \sum_1^k w_k \delta(D_k, C_j) \tag{20}$$

Chapter 5: Experiments, Results and Analysis

5.1   Final Proposed Scheme

Figure 17 illustrates a simplified block diagram of our proposed gait recognition system. A decision-level fusion scheme takes as inputs the decision labels from up to nine (9) different gait recognition sub-systems. The sub-systems are designed in such a way to utilize three (3) different gait representations (pre-processing techniques) with three (3) different matching methods.
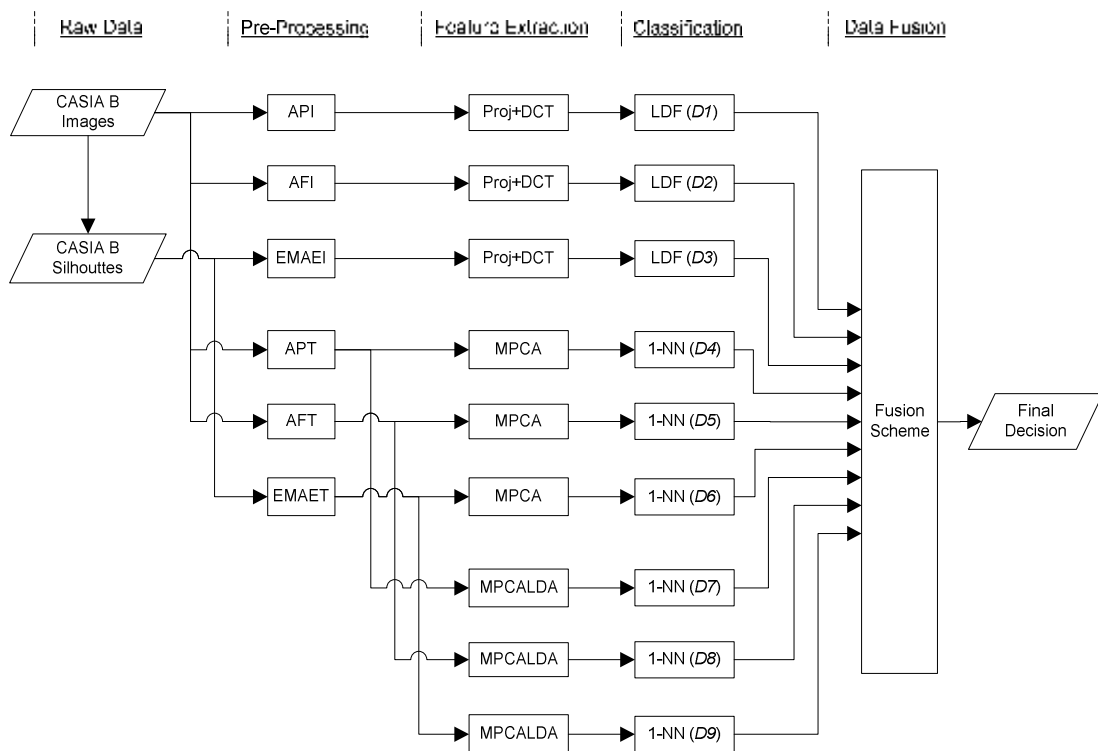
Figure 17: Block Diagram of the proposed Gait Recognition System

System is made-up of the following components:

Gait Representation (Pre-processing) Techniques:

1. Accumulated Prediction Image (API)
2. Accumulated Flow Image (AFI)
3. Edge-Masked Active Energy Image (EMAEI)

Each of the above-listed gait images is used, as the feature pool, three times; implementing the following methods:

1. Image Projection (Proj), followed by a linear classifier (LDF). We shall label this as Method 1 or (*Mthd1*).

2. MPCA, followed by Nearest Neighbor classifier (1-NN). We shall label this as Method 2 or (*Mthd2*).

3. MPCALDA, followed by Nearest Neighbor classifier (1-NN). We shall label this as Method 3 or (*Mthd3*).

Note that inputs to both Mthd2 and Mthd3 are the tensorial gait representations, APT, AFT and EMAET.

5.2   Testing Methodology

We have carried-out approximately 300 different experiments using different combinations of feature extraction and classification methods including those covered in this report and a few others such as; 2D Discrete Cosine Transform (DCT), Dynamic Time Warping (DTW) and Procrustes Shape Analysis. Amongst these experiments we show here only the ones contributing to the final system. The next section lists-up the four different experimental setups used to evaluate the proposed method. In each setup we carry-out 11 experimental groups. Each group comprises number of experiments equivalent to the number of testing probes as detailed for each case. The groups are as follows:

Single-level Gait Recognition:

1. API + *Mthd1*, labeled Recognizer 1 or (*R1*)
2. AFI + *Mthd1*, labeled Recognizer 2 or (*R2*)
3. EMAEI + *Mthd1*, labeled Recognizer 3 or (*R3*)
4. APT + *Mthd2*, labeled Recognizer 4 or (*R4*)
5. AFT + *Mthd2*, labeled Recognizer 5 or (*R5*)
6. EMAET + *Mthd2*, labeled Recognizer 6 or (*R6*)
7. APT + *Mthd3*, labeled Recognizer 7 or (*R7*)
8. AFT + *Mthd3*, labeled Recognizer 8 or (*R8*)
9. EMAET + *Mthd3*, labeled Recognizer 9 or (*R9*)

Decision-Level Fusion Schemes:

10. Three-decision Fusion Scheme as follows:

    - *3Fus123*: 3-decision fusion of *R1*, *R2* and *R3*.
    - *3Fus456*: 3-decision fusion of *R4*, *R5* and *R6*.
    - *3Fus789*: 3-decision fusion of *R7*, *R8* and *R9*.
    - *3Fus147*: 3-decision fusion of *R1*, *R4* and *R7*
    - *3Fus258*: 3-decision fusion of *R2*, *R5* and *R8*
    - *3Fus369*: 3-decision fusion of *R3*, *R6* and *R9*

    Each of these fusion experiments is repeated twice, using Unweighted Voting (UWV) and Weighted Voting (WV) techniques.

11. Nine-decision Fusion (*9Fus*) Scheme. Similarly, this is repeated twice using the UWV and WV techniques.

5.3 Experimental Setups

As mentioned earlier, and for the sake of evaluation and analysis of our proposed method, we typically need two things:

1. Gait database; suitable to the application under study. In our case, CASIA B database was a good choice for the reasons listed in section 2.1.1. It was captured indoor, which makes it suitable to our experiments that have been designed to test against change of carrying conditions and clothing, with the environment (or background) being controlled. Once results are obtained, and as we shall see they are promising ones, we can take the model outdoor and test it using different databases and sets of experiments. This is intended for future work.

2. Experimental setups; which means dividing our database into training (gallery) data and testing (probe) data. Number of sequences in each of the two data groups has been varied based on similar setups taken from six (6) recently published papers that studied the same topic and used the same database. It is worth mentioning that these papers have reported the highest recognition rates in the literature up until the date of writing this report. Total number of sequences is obviously the same in all the 4 setups and these are:

a. Six sequences for normal-condition walking, labeled *nm1*, *nm2*, *nm3*, nm4, *nm5* and *nm6*

b. Two sequences for walking carrying bag, labeled *bg1* and *bg2*.

c. Two sequences for walking wearing coat (cloth), labeled *cl1* and *cl2*.

The data breakdown in each of the setups is as follows:

a. Setup 1 [13]:

- Training Data (*Tr1*): 3 normal sequences; *nm1*, *nm2* and *nm3*

- Testing Data : 3 sets for 3 experiments;
  - *Ts11*: 3 normal sequences; *nm4*, *nm5* and *nm6*
  - *Ts12*: 2 bag sequences; *bg1* and *bg2*
  - *Ts13*: 2 coat sequences; *cl1* and *cl2*

  This setup was the primary one, used extensively in our experiments, and the results reported in [13] were our primary target. Therefore, most of the experiments carried out, especially on the API gait representation and the subsequent versions using different feature extraction and/or classification methods, have actually used this paper, and its setup and results, as reference. The paper used the Active Energy Image (AEI) method for gait representation, followed by two-dimensional Locality Preserving Projection (2DLPP) for dimensionality reduction and feature extraction, and finally Nearest Neighbor classifier.

b. Setup 2 [14, 34, 66]:

- Training Data (*Tr2*): 4 normal sequences; *nm1*, *nm2*, *nm3* and *nm4*

- Testing Data : 3 sets for 3 experiments;
  - *Ts21*: 2 normal sequences; *nm5* and *nm6*
  - *Ts22*: 2 bag sequences; *cl1* and *cl2*
  - *Ts23*: 2 coat sequences; *bg1* and *bg2*

The work in [14] implements feature-level fusion of two gait signatures; Shifted Energy Image (SEI) and Gait Structural Profile (GSP). Paper [34] implements the Random Forest algorithm for feature ranking, and produced a Masked GEI, followed by Canonical Discriminant Analysis (CDA). Gait Pal and Pal Entropy (GPPE) Image is used in [66] as the gait

representation, followed by PCA and SVM classifier. Experiments in [66] were carried out on 98 subjects only.

c. Setup 3 [16]. This paper implements a temporal gait template, named Chrono-Gait Image (CGI), followed by PCA+LDA for feature extraction and 1-NN classifier.

- Training Data (*Tr3*): 5 normal sequences; *nm1*, *nm2*, *nm3*, *nm4* & *nm5*.
- Testing Data : 5 sets for 5 experiments;
  - *Ts31*: normal sequence *nm6*
  - *Ts32*: bag sequence *bg1*
  - *Ts33*: bag sequence *bg2*
  - *Ts34*: coat sequence *cl1*
  - *Ts35*: coat sequence *cl2*

d. Setup 4 [15]. This paper uses the concept of 'Axis of Least Inertia' to produce gait features.

- Training Data (*Tr4*): 3 normal sequences; *nm1*, *nm2* and *nm3* plus 1 bag sequence *bg1* and 1 coat sequence *cl1*.
- Testing Data : 5 sets for 5 experiments;
  - *Ts41*: 1 normal sequence; *nm4*
  - *Ts42*: 1 normal sequence; *nm5*
  - *Ts43*: 1 normal sequence; *nm6*
  - *Ts44*: 1 bag sequence; *bg2*
  - *Ts45*: 1 coat sequence; *cl2*

5.4   Results and Analysis

5.4.1 Results using experimental Setup 1

Table 4 summarizes the recognition rates obtained from the experiments carried out using Training Data *Tr1* and Testing Data *Ts11*, *Ts12* and *Ts13*, and Figure 18 represents them graphically. Note that in this arrangement, sequences of *bag* and *coat* are unseen during training. We compare our results with those obtained by Zhang et al. in [13].

Table 4: Recognition Rates (%) of experiments using Setup 1 [13]

	R1	R2	R3	R4	R5	R6	R7	R8	R9	9Fus (UWV)	9Fus (WV)	Zhang et al. [13]
Ts11	**98.66**	98.39	65.59	93.55	93.28	90.59	89.25	76.61	90.32	**99.19**	**99.19**	98.40
Ts12	87.90	84.68	41.53	75.81	82.26	52.02	68.95	53.23	44.76	**96.77**	95.97	91.95
Ts13	40.32	51.21	42.74	42.34	51.21	**81.05**	39.92	41.94	76.61	83.87	**88.31**	72.19
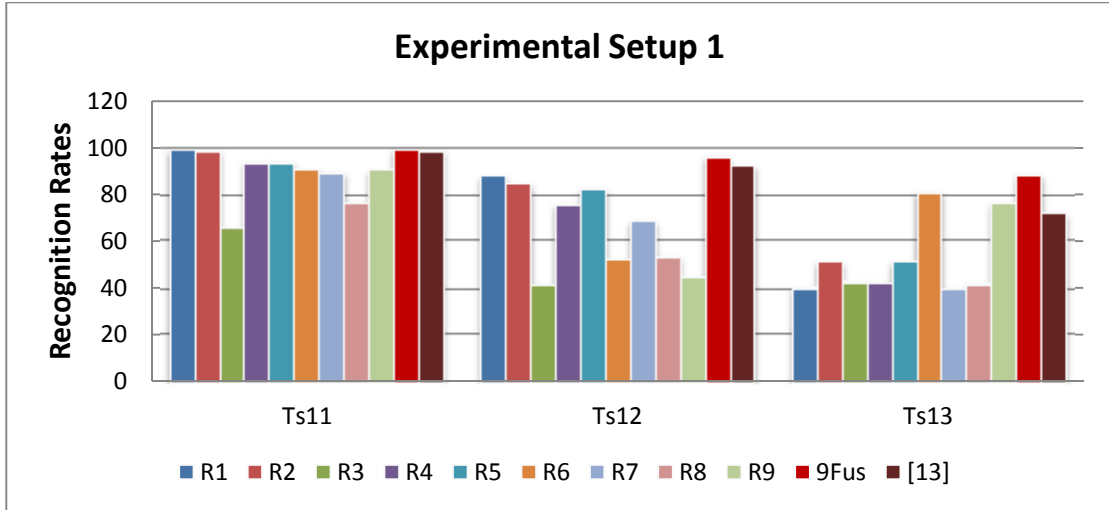
Figure 18: Recognition Rates (%) of the experiments groups *Ts11 (nm1, nm2 & nm3), Ts12 (bg1 & bg2)* and *Ts13 (cl1 & cl2)*

We notice that it is possible to achieve high recognition rates when testing *normal* sequences (*Ts11*) even without fusion, using a simple linear classifier, as in *R1* and *R2*. This would be observed in all the succeeding results of other setups, when testing for *normal* sequences, which would prove that our gait representation techniques, especially the API, are efficient when covariates are controlled.

We notice also that experiments using API with Projection and LDF were good enough to produce very good recognition rates when testing for *bag* sequences (*Ts12*). They are yet not good enough compared to those reported in [13]. 9-decision fusion (*9Fus*) method, however, outperforms Zhang et al. by almost 5%. EMAEI is still not efficient representation for the two cases of *normal* and *bag* sequences, especially in *R3*. However, it works well in the case of *coat* sequences (*Ts13*), even without optimization. We notice that the two experiments using *R6* (MPCA) and *R9* (MPCALDA) have scored Recognition Rates that outperform the target by approximately 9% and 4% respectively. 9-decision fusion has even improved this to score approximately 15% above target.

It is noteworthy here that this setup is the most onerous compared to the three following it. We'll see in the next setups that additional sequences are presented during training. In Setup 4, we even have *bag* and *coat* sequences presented during training. These arrangements are expected to improve the average recognition rates of the base Recognizers and the overall fusion results.

We show herein a comparison between Unweighted and Weighted fusion techniques at the two levels; 3-decision fusion and 9-decision fusion. In order to do that, we had first to estimate the base classifiers' accuracies and assign them authority values, and eventually weights. For this purpose, we have split the training data used in Setup 1 (*Tr1*) into two parts; 2 *normal* sequences (*nm1* and *nm2*) for training and 1 *normal* sequence (*nm3*) to test for the classifiers' accuracy. This gives us a general indication on each classifier's relative performance. Table 5 lists the errors (e_k), authority values (a_k) and weights (w_k) given to each of the nine base recognizers using the WV method (equation 19). These weights are used throughout the four experimental setups.

Table 5: Weights given to base recognition modules

	R1	R2	R3	R4	R5	R6	R7	R8	R9
e_k	0.081	0.057	0.629	0.081	0.089	0.105	0.177	0.210	0.032
a_k (WV)	0.919	0.944	0.371	0.919	0.911	0.895	0.823	0.790	0.968
w_k (WV)	**0.122**	**0.125**	**0.049**	**0.122**	**0.121**	**0.119**	**0.109**	**0.105**	**0.128**

Note that the weights shall sum to 1. Now, we'll use these weights in the voting scheme for fusion and we'll compare the results obtained from the presented fusion techniques. After applying this to the experiments in Setup 1, we get the fusion Recognition Rates (%) as in Table 6 and Figure 19. Note that schemes F1-F6 combine recognizers of the same [feature extraction + classifier] pair, while varying gait representation techniques; whereas schemes F1*-F6* combine recognizers of the same representation techniques while varying [feature extraction + classifier] pairs. Based on the fusion experiments, we conclude the following:

- In the case of 3-decision fusion; fusing decisions of different representations using the same [feature extraction + classifier] pair has performed better than fusing decisions from different classifiers. This will be picked in the succeeding experimental setups.

- The WV method has performed better than the Unweighted (UWV) method in most of the cases.

Table 6: Recognition Rates (%) of fusion techniques using Setup 1 [13]

Fusion Scheme	Ts11	Ts12	Ts13
3Fus123/ (UWV) – F1	**99.1935**	87.9032	52.0161
3Fus123 / (WV) – F2	**99.1935**	89.5161	57.2581
3Fus456 / (UWV) – F3	96.5054	81.0484	66.9355
3Fus456 / (WV) – F4	97.0430	81.0484	58.4677
3Fus789 / (UWV) – F5	93.2796	70.1613	63.7097
3Fus789 / (WV) – F6	96.2366	62.9032	**79.8387**
3Fus147/ (UWV) – F1*	97.5806	85.0806	42.3387
3Fus147 / (WV) – F2*	97.5806	85.0806	45.9677
3Fus258 / (UWV) – F3*	96.7742	84.2742	56.4516
3Fus258 / (WV) – F4*	98.1183	89.1129	56.4516
3Fus369 / (UWV) – F5*	92.4731	52.0161	64.9194
3Fus369 / (WV) – F6*	93.0108	52.4194	57.6613
9Fus / (UWV) – F7	99.1935	**96.7742**	83.8710
9Fus / (WV) – F8	**99.1935**	95.9677	**88.3065**
Zhang *et al.* [13]	98.4000	91.9500	72.1900
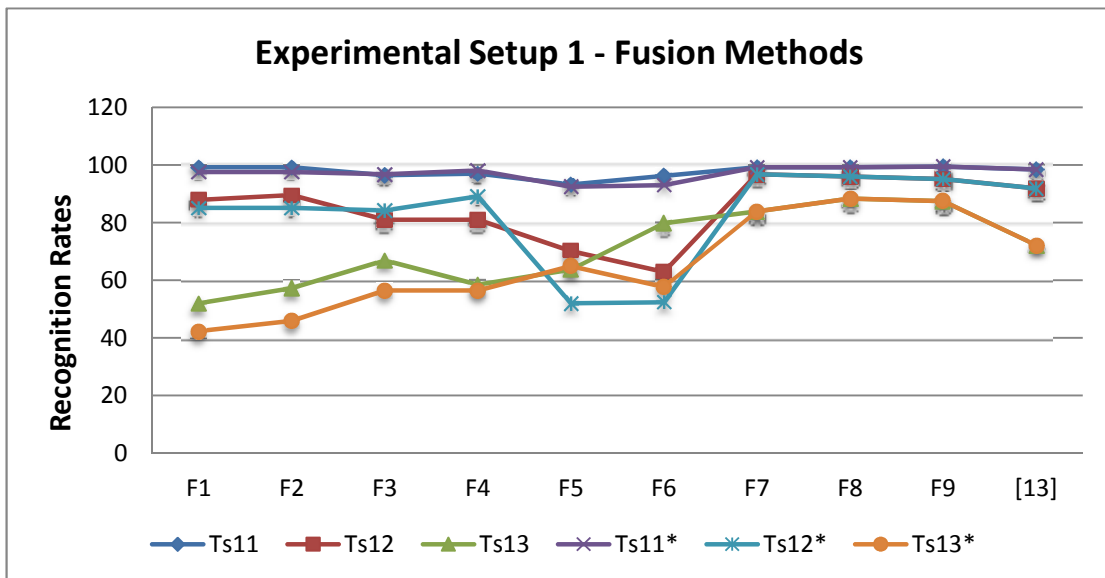
Figure 19: Recognition Rates (%) of the experiments groups *Ts11* (*nm1*, *nm2 & nm3), Ts12 (bg1 & bg2)* and *Ts13 (cl1 & cl2),* using different methods of fusion. *Ts11*, *Ts12* and *Ts13* are the experimental results for fusion schemes *F1*-*F6**

- In the 9-decision fusion schemes (*9Fus*), each of the two tested schemes (unweighted and weighted) has outperformed the published results. Again, the WV would average the best results.

5.4.2 Results using experimental Setup 2

Table 7 summarizes the recognition rates obtained for the experiments carried out using Training Data *Tr2* and Testing Data *Ts21*, *Ts22* and *Ts23*, and Figure 20 represents them graphically. Sequences with *bag* and *coat* are still unseen during training. The difference here is one additional normal sequence (*nm4*) in the training data. We compare our results with those obtained by Huang et al. in [14], Dupuis et al. in [34] and Jeevan et al. [66]

Table 7: Recognition Rates (%) of experiments using Setup 2 [14, 34, 66]

	R1	R2	R3	R4	R5	R6	R7	R8	R9	9Fus (UWV)	9Fus (WV)	Huang et al. [14]	Dupuis et al. [34]	Jeevan et al. [66]*
Ts21	99.19	**99.60**	77.82	94.76	95.57	91.13	94.36	93.95	96.77	**99.60**	**99.60**	99	98.79	93.36
Ts22	43.15	54.03	50.00	46.77	54.03	85.08	37.50	46.77	**88.71**	87.50	**91.93**	64	**92.74**	22.44
Ts23	**88.71**	**88.71**	51.21	79.03	84.27	53.23	74.19	72.18	60.48	**98.39**	97.58	72	77.82	56.12

*Experiments carried out on 98 subjects only

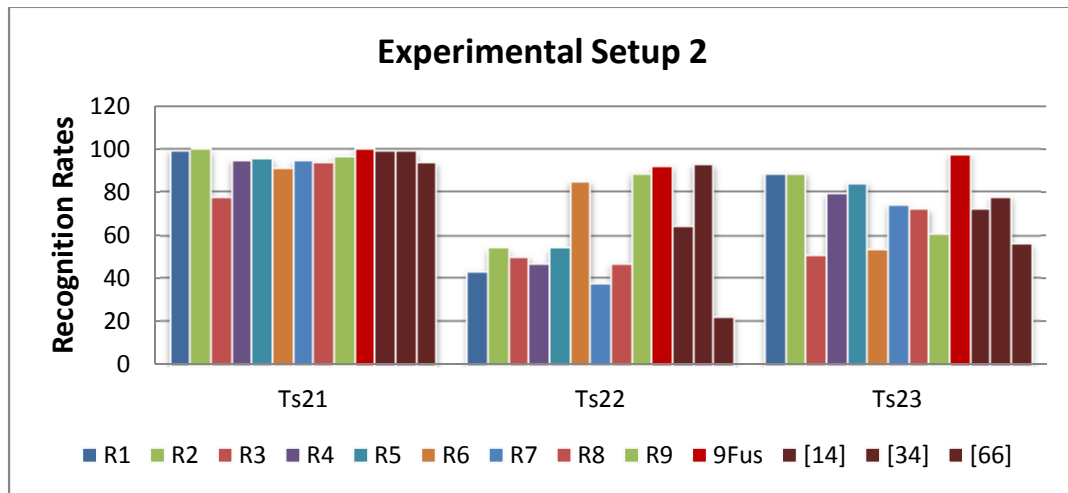Figure 20: Recognition Rates (%) of the experiments groups *Ts21* (*nm1*, *nm2*), *Ts22 (cl1 & cl2)* and *Ts23 (bg1 & bg2)*

It is still noticeable that high recognition rates are possible without fusion in the two cases of *R1* and *R2*, this time for both cases of *normal* sequences (*Ts21*) and *bag* sequences (*Ts23*) which, along with the result of *9Fus*, exceed those obtained in the three reported papers.

Again, EMAEI gait representation has worked well for *coat* sequences (*Ts22*), especially with the MPCALDA algorithm *R9*. The rate in [34] still exceeds this one by more than 4%. But it is higher than the rate obtained in *9Fus (WV)* by less than 1%. It is worth mentioning here that Dupuis et al. have adopted a feature selection technique yielding some sort of masked gait images. Their method seems to have worked well mainly in the case of *coat* sequences. And we recall that our EMAEI is un-optimized and currently based on human observation only. Better results are anticipated upon optimization of this gait representation method.

It can be easily noticed that the individual rates of the base Recognizers, especially for *normal* sequences, have averaged higher compared to those in Setup 1. This is expected having introduced one additional sequence in training. General behavior of all base methods is also similar to that in Setup 1. As for *Ts23* (*bag* sequences) Experiments using API with Projection and LDF were good enough to outperform those reported in the target papers. 9-decision fusion method results in even better rate that is almost 10% higher compared to [34].

Table 8 shows the fusion results using Setup 2. It can be seen that 3-decision fusion techniques are more effective here compared to the more onerous Setup 1. We still use the same accuracy weights estimated in Setup 1 for simplification, although it is possible to add more validation sequences to optimize the weights.

Table 8: Recognition Rates (%) of fusion techniques using Setup 2 [14, 34, 66]

	3Fus123 (UWV)	3Fus123 (WV)	3Fus456 (UWV)	3Fus456 (WV)	3Fus789 (UWV)	3Fus789 (WV)	9Fus (UWV)	9Fus (WV)	Huang et al. [14]	Dupuis et al. [34]	Jeevan et al. [66]*
Ts21	**99.60**	**99.60**	97.98	97.18	98.79	98.79	**99.60**	**99.60**	99	98.79	93.36
Ts22	57.26	60.89	64.52	62.50	69.35	89.52	87.50	**91.93**	64	92.74	22.44
Ts23	90.60	**91.94**	82.66	85.08	79.44	78.63	**98.39**	97.58	72	77.82	56.12

5.4.3 Results using experimental Setup 3

Table 9 summarizes the recognition rates obtained for the experiments carried out using Training Data *Tr3* and Testing Data *Ts31*, *Ts32*, *Ts33, Ts34* and *Ts35*, and Figure 21 represents them graphically. Note that we have added again one more normal sequence (*nm5*) to the training data, whereas sequences with *bag* and *coat* are still unseen during training. We compare our results with those obtained by Wang et al. in [16].

Table 9: Recognition Rates (%) of experiments using Setup 3 [16]

	R1	R2	R3	R4	R5	R6	R7	R8	R9	9Fus (UWV)	9Fus (WV)	Wang et al. [16]
Ts31	99.19	**100**	78.23	96.77	97.58	93.55	96.77	98.39	98.39	**100**	**100**	100
Ts32	**89.52**	87.90	52.42	79.03	82.26	55.65	79.03	82.26	68.55	**99.19**	97.58	68.52
Ts33	**91.94**	90.32	58.06	83.07	85.48	57.26	78.23	77.42	61.29	**99.19**	98.39	75.00
Ts34	41.94	54.84	50.81	41.94	50.81	87.90	38.71	46.77	**92.74**	90.32	90.32	49.07
Ts35	44.35	60.48	54.84	54.84	62.10	88.71	46.77	53.23	**94.36**	91.94	91.94	44.44
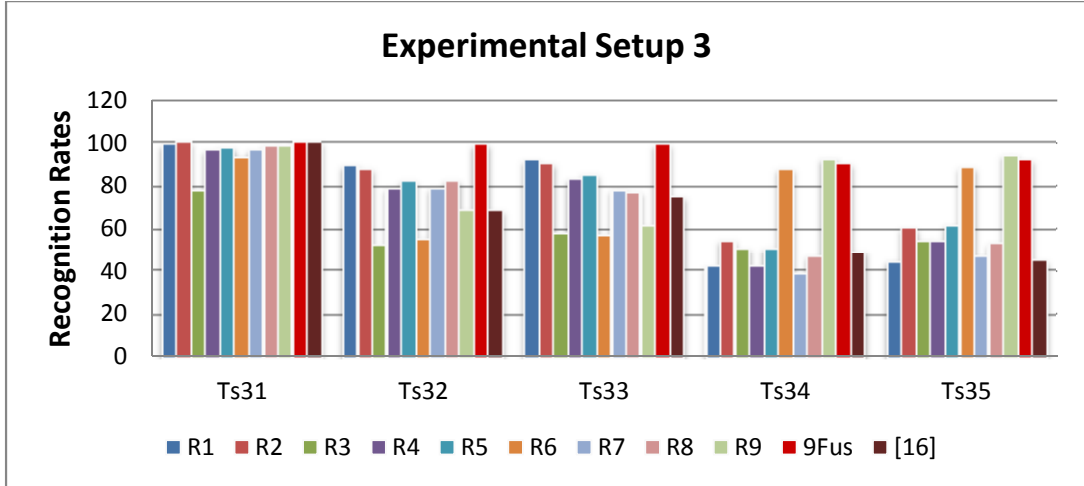
Figure 21: Recognition Rates (%) of the experiments groups *Ts31 (nm6), Ts32 (bg1), Ts33 (bg2), Ts34 (cl1)* and *Ts35 (cl2)*

Most of the base methods have performed extremely well. It can be noticed that one of these methods, *R2* has outperformed the target in all the five testing experiments. Fusion has improved this dramatically. We can see in the case of *coat* (*Ts34* and *Ts35*) our fusion rates are approximately 50% higher than those reported in [16]. Table 10 shows the fusion results using Setup 3. It can be seen that all 3-decision fusion techniques outperforms the target results and score well. As expected, 9-decision fusion improves this further.

Table 10: Recognition Rates (%) of fusion techniques using Setup 3 [16]

	3Fus123 (UWV)	3Fus123 (WV)	3Fus456 (UWV)	3Fus456 (WV)	3Fus789 (UWV)	3Fus789 (WV)	9Fus (UWV)	9Fus (WV)	Wang et al. [16]
Ts31	99.19	**100**	98.39	98.39	**100**	100	**100**	**100**	100
Ts32	90.32	**91.94**	82.26	85.48	83.87	83.87	**99.19**	97.58	68.52
Ts33	91.94	**95.16**	82.26	86.29	86.29	82.26	**99.19**	98.39	75.00
Ts34	60.48	60.48	71.77	62.10	73.39	**90.32**	90.32	90.32	49.07
Ts35	63.71	65.32	75.81	71.77	74.19	**94.35**	91.94	91.94	44.44

5.4.4 Results using experimental Setup 4

Table 11 summarizes the recognition rates obtained for the experiments listed in section 5.2, using Training Data *Tr4* and Testing Data *Ts41*, *Ts42*, *Ts43, Ts44* and *Ts35*, and Figure 22 represents them graphically. This is the only setup in which training data has samples from all three groups, *normal*, *bag* and *coat*. It is expected that this arrangement should enhance the final overall recognition rate, especially when testing for *bag* and *coat* sequences. We compare our results with those obtained by Kumar et al. in [15].

Table 11: Recognition Rates (%) of experiments using Setup 4 [15]

	R1	R2	R3	R4	R5	R6	R7	R8	R9	9Fus (UWV)	9Fus (WV)	Kumar et al. [15]
Ts41	96.77	**98.39**	66.94	94.36	93.55	95.97	90.32	94.35	**98.39**	**99.19**	**99.19**	93.33
Ts42	**100**	99.19	58.06	95.97	92.74	95.16	87.90	90.32	98.39	**100**	**100**	96.66
Ts43	**99.19**	96.77	66.94	94.36	93.55	91.94	87.10	91.94	97.58	**100**	**100**	92.50
Ts44	98.39	**99.19**	63.71	93.55	90.32	87.10	84.68	87.90	93.55	**100**	**100**	76.66
Ts45	85.48	95.16	56.45	84.68	85.48	95.97	70.97	77.42	**97.58**	99.19	**100**	79.16
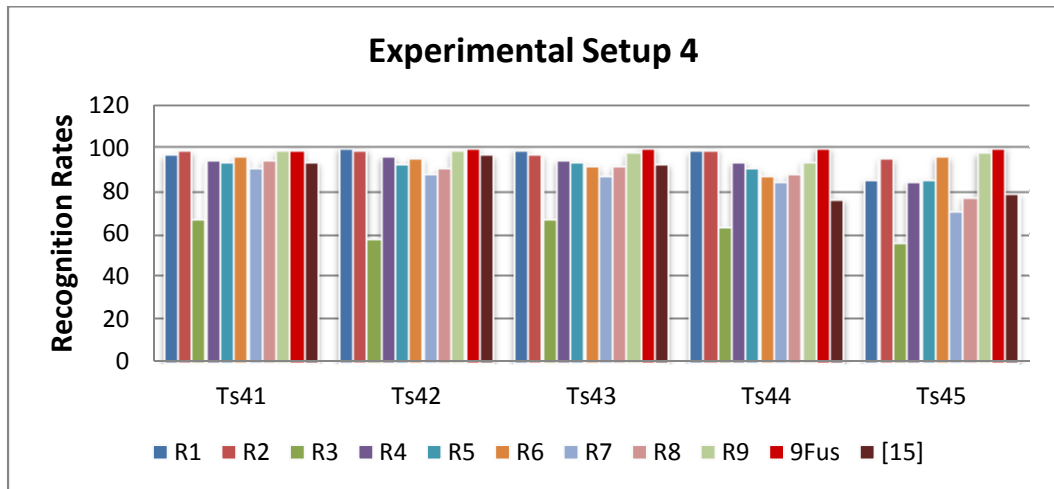
Figure 22: Recognition Rates (%) of the experiments groups *Ts41 (nm4), Ts42 (nm5), Ts43 (nm6), Ts44 (bg2)* and *Ts45 (cl2)*

Exactly as expected, recognition rates have topped high due to training for the two variables of carrying condition (*bag*) and clothing (*coat*). Most of our base methods, especially *R1, R2* and *R9,* have outperformed those in [15]. This is especially noticeable in the two cases of *Ts44* (*bg2*) and *Ts45* (*cl2*). EMAEI is apparently not so efficient when combined with Image Projection and LDF (*R3*). However, it performs amazingly well with MPCALDA (*R9*). Table 12 shows the

fusion results using Setup 4. 3-decision fusion techniques also perform well, and they all outperform the target results. Almost all of the Recognition Rates achieved by 9-decision fusion have hit 100%.

Table 12: Recognition Rates (%) of fusion techniques using Setup 4 [15]

	3Fus123 (UWV)	3Fus123 (WV)	3Fus456 (UWV)	3Fus456 (WV)	3Fus789 (UWV)	3Fus789 (WV)	9Fus (UWV)	9Fus (WV)	Kumar et al. [15]
Ts41	**99.19**	**99.19**	97.58	98.39	95.97	97.58	**99.19**	**99.19**	93.33
Ts42	**99.19**	**99.19**	97.58	97.58	96.77	98.39	**100**	**100**	96.66
Ts43	97.58	96.77	**98.39**	**98.39**	97.58	97.58	**100**	**100**	92.50
Ts44	**99.19**	**99.19**	96.77	97.58	96.77	97.58	**100**	**100**	76.66
Ts45	92.74	95.97	95.97	95.97	93.55	**99.19**	99.19	**100**	79.16

In general, we notice that the unweighted and weighted 9-decision voting schemes outperform all the recently published methods that used the state-of-the-art algorithms. While this is the case, it may not be always needed. In most of the tests presented above, two of the base methods, *R1* and *R2*, have also individually exceeded the target results. To recall, these methods use the API and AFI gait representations respectively, using the simple image projection for feature extraction and linear classifier. The EMAEI gait representation, although still require improvement and optimization, has performed well for the *coat* sequences, particularly when combined with the MPCALDA algorithm (*R9*). As can be noticed, experimental setups gradually decrease in difficulty moving from Setup 1 to Setup 4. 3-decision fusion techniques perform amazingly well in Setup 4. They also outperform the target results in Setup 3, and with less percentage for Setup 2. In Setup 1, however, we need 9-decision fusion to achieve our targets. It should be also noted that the fusion results in Setups 2-4 were achieved using the weights from Setup 1. These might be optimized further, and results improved, by utilizing the additional training data for validation.

As a final test, and based on the findings of our experiments above, we'll evaluate another 3-decision fusion combination using base recognizers *R1*, *R2* and *R9* which yielded the best average individual recognition rates in the four setups. We'll also concentrate on the Weighted Voting (WV) scheme only. The method is labeled 3Fus129. Tables 13-16 below list these final results and compare them to the 9-decision fusion rates and the targets. We can see this 3-decision fusion method

outperforms almost all the target results in the four setups. Obviously, the 9-decision fusion enhances the rates further.

Table 13: Recognition Rates (%) from fusion of *R1*, *R2* and *R9* using Setup 1 [13]

	3Fus129 (WV)	9Fus (WV)	Zhang et al. [13]
Ts11	**98.93**	**99.19**	98.40
Ts12	86.29	**95.97**	91.95
Ts13	**79.84**	**88.31**	72.19

Table 14: Recognition Rates (%) from fusion of *R1*, *R2* and *R9* using Setup 2 [14, 34, 66]

	3Fus129 (WV)	9Fus (WV)	Huang et al. [14]	Dupuis et al. [34]	Jeevan et al. [66]
Ts21	**99.60**	**99.60**	99	98.79	93.36
Ts22	86.69	91.93	64	**92.74**	22.44
Ts23	**92.73**	**97.58**	72	77.82	56.12

Table 15: Recognition Rates (%) from fusion of *R1*, *R2* and *R9* using Setup 3 [16]

	3Fus129 (WV)	9Fus (WV)	Wang et al. [16]
Ts31	**100**	**100**	100
Ts32	**91.94**	**97.58**	68.52
Ts33	**95.97**	**98.39**	75.00
Ts34	**90.32**	**90.32**	49.07
Ts35	**91.94**	**91.94**	44.44

Table 16: Recognition Rates (%) from fusion *R1*, *R2* and *R9* using Setup 4 [15]

	3Fus129 (WV)	9Fus (WV)	Kumar et al. [15]
Ts41	**100**	**99.19**	93.33
Ts42	**99.19**	**100**	96.66
Ts43	**99.19**	**100**	92.50
Ts44	**99.19**	**100**	76.66
Ts45	**97.58**	**100**	79.16

Chapter 6: Conclusions and Future Works

In this thesis, gait recognition was studied in details, covering the different gait databases and gait approaches. It is noticed that the problem has become the research topic in many of the recent biometrics and computer vision research papers. This is an indication of the attractiveness and potentiality of gait as a biometric signature. Major part of the research work was dedicated to overcome and solve for the many covariates that impact the gait quality and robustness, such as carrying conditions and clothing. Our study has primarily focused on these two factors. Additional tests are recommended in the future to study the impact of and the system's performance against other variables, especially the viewing angle. In the outcome of our comprehensive literature review, and based on the results of the extensive experiments carried out, it is concluded that model-free gait approaches, and particularly spatio-temporal (accumulated error) and energy-based methods, can perform well and should be focused on for the future gait recognition research.

Nine different gait approaches have been evaluated for individual performance, and when combined in fusion schemes. Three gait representation methods are used; Accumulated Prediction Image (API) and two novel representation techniques namely, Accumulated Flow Image (AFI) and Edge-Masked Active Energy Image (EMAEI). It was shown that the first two methods are effective in producing distinctive features. The EMAEI method, although requires further optimization, has performed well in the case of *coat* sequences as it removes the parts of the body that correspond to static features. Therefore, it contributes efficiently in the final fusion scheme. We recommend that the mask-creating method is automated using some techniques like filtering, wrapping, decision trees or similar feature selection techniques.

The second step was feature extraction. Three methods have been evaluated. One simple image projection with 1D DCT, and two multilinear techniques namely MPCA and MPCALDA. This was an effective combination of simplicity and compatibility. The two multilinear techniques are suitable to the tensorial nature of the gait features, being 3-mode data. The advantage was to keep data in its raw nature. Each of the three gait representation images was tested with each of the feature extraction methods, yielding a model of nine different base recognizers. It is

noteworthy that we used two of the simplest classifiers, Linear Discriminant Functions (LDF) and 1 Nearest Neighbor classifiers for the base decisions. As part of future work, more advanced classifiers; such as Polynomial Networks, Support Vector machines (SVM) and Neural Networks might be tested.

The last part of the thesis was dedicated to analyze the proposed method's performance, taken from the results of the experiments. Major part of that was for testing and evaluating the different decision-level fusion techniques. Fusion has been adopted recently in multi-modal biometric systems, and has shown promising results in solving for issues like, robustness and sensors' sensitivity. In addition, it is shown that fusion can combine the base classifiers' discriminating power and boost the overall system's performance. Decision-level fusion is simple and straightforward compared to feature-level and score-level fusion techniques. It is concluded that fusion is recommended in modern biometric systems, particularly the behavioral ones like gait. Based on a comparison between unweighted and weighted voting fusion schemes, it was found that, weighting can improve the performance. As for the two voting schemes tested in this work; Unweighted Voting (UWV) and Weighted Voting (WV), the WV method has shown improvements in most of the fusion tests carried out. Further evaluation of the fusion schemes might be carried out as part of the future work, and especially using different databases to ensure the method is insensitive to the data presented, and can generalize well.

References

[1]   J. Little and J. Boyd. "Biometric Gait Recognition," in *Advanced Studies in Biometrics*, *LNCS* 3161, pp. 19–42, 2005

[2]   S. Niyogi, E. Adelson. "Analyzing and Recognizing Walking Figures in XYT," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, Wash, USA,* pp. 469–474, 1994

[3]   T. Shanableh, K. Assaleh, L. Al-Hajjaj and A. Kabani. "Gait Recognition System Tailored for Arab Costume of The Gulf Region," in *Proc. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT),* Ajman, UAE, pp. 544-549, 2009

[4]   H. Lu, K. Plataniotis, and A. Venetsanopoulos. "MPCA: Multilinear Principal Component Analysis of tensor Objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18-39, Jan 2008.

[5]   M. Nixon, J. Carter. "Advances in Automatic Gait Recognition," in *Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition,* Seoul, Korea, pp. 139–144, 2004

[6]   M. Nixon, and J. Carter. "Automatic Recognition by Gait," in *Proc. the IEEE, v*ol. 94, no. 11, pp. 2013-2024, Nov 2006

[7]   G. Johansson. "Visual perception of biological motion and a model for its analysis," in *Perception and Psychophysics,* vol. 14, no. 2, pp 201–211, 1973

[8]   M. Murray, A. Drought, and R. Kory. "Walking patterns of normal men," *The Journal of Bone and Joint Surgery,* vol. 46, no. 2, pp. 335–360, 1964.

[9]   B. Bhanu, J. Han. "Bayesian-based performance prediction for gait recognition," in *Proc. IEEE Workshop on Motion and Video Computing, Orlando, Florida,* pp 145–150, 2002

[10]   J. Laszlo, M. van de Panne, E. Fiume. "Limit cycle control and its application to the animation of balancing and walking," in *Proc. the 23rd annual conference on Computer graphics and interactive techniques (SIGGRAPH).* New Orleans, LA, USA, pp 155–162, 1996

[11]   S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. "The HumanID gait challenge problem: Data sets, performance and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 2, pp. 162–177, Feb. 2005.

[12]   V. Von Tscharner, B. Goepfert. "Gender dependent EMGs of runners resolved by time/frequency and principal pattern analysis," *Journal of*

Electromyography and Kinesiology, vol.13, no. 3, pp 253–272, 2003

[13]   E. Zhang,Y. Zhao, W. Xiong. "Active energy image plus 2DLPP for gait recognition," *Journal of Signal Processing*, vol. 90, no. 7, pp. 2295–2302, 2010

[14]   X. Huang, N. Boulgouris. "Gait Recognition With Shifted Energy Image & Structural Feature Extraction," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2256-2268, 2012

[15]   H. Kumar, H. Nagendraswamy. "Gait Recognition- An Approach Based on Interval Valued Features," in *Proc. International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, pp. 1-6, Jan. 04-06, 2013

[16]   C. Wang, J. Zhang, L. Wang, J. Pu, X. Yuan. "Human Identification Using Temporal Information Preserving Gait Template," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, Nov. 2012

[17]   D. Gafurov. "Security Analysis of Impostor Attempts with Respect to Gender in Gait Biometrics," in *Proc. First IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, Crystal City, VA, USA, pp. 1-6, 2007.

[18]   D. Gafurov, E. Snekkenes*,* and P. Bours. "Spoof Attacks on Gait Authentication System," *IEEE Transactions on Information Forensics and Security,* vol. 2, no. 3, pp. 491-502, Sep 2007

[19]   S. Lee, Y. Liu, and R. Collins. "Shape Variation-Based Frieze Pattern for Robust Gait Recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Minneapolis, MN, USA, pp. 1-8, 2007

[20]   Z. Liu, and S. Sarkar. "Outdoor recognition at a distance by fusing gait and face," *Image and Vision Computing Journal*, vol. 25, no. 6, pp. 817-832, 2007

[21]   C. BenAbdelkader, R. Cutler, and L. Davis. "Gait Recognition Using Image Self-Similarity," *EURASIP Journal on Applied Signal Processing*, vol. 4, pp. 572–585, 2004

[22]   J. Hayfron-Acquah, M. Nixon, and J. Carter. "Automatic gait recognition by symmetry analysis," *Pattern Recognition Letters,* vol. 24, no. 13, pp. 2175–2183, 2003.

[23]   J. Little, and J. Boyd. "Recognizing People by Their Gait: The Shape of Motion," *Videre,* vol. 1, no. 2, pp. 1-33, 1998.

[24]   A. Kale, A. Rajagopalan, A. Sundaresan, N. Cuntoor, A. Roy-Chowdhury, V. Kruger, and R. Chellappa. "Identification of humans using gait," *IEEE Transactions on Image Processing,* vol. 13, no. 9, pp. 1163–1173, Sep. 2004.

[25] A. Sundaresan, A. Roy-Chowdhury, and R. Chellappa. "A hidden Markov model based framework for recognition of humans from gait sequences," in *Proc. IEEE International Conference on Image Processing (ICIP), Barcelona, Spain,* vol. 2, pp. 96-96, 2003

[26] R. Collins, R. Gross, and J. Shi. "Silhouette-based human identification from body shape and gait," in *Proc. Fifth IEEE International Conference on Face and Gesture Recognition*, Washington, DC, USA, pp. 366–371, 2002

[27] R. Zhang, C. Vogler, and D. Metaxas. "Human gait recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Washington, DC, USA, vol. 1, pp. 18, 2004

[28] L. Wang, T. N. Tan, W. M. Hu, and H. Z. Ning. "Automatic gait recognition based on statistical shape analysis," *IEEE Transactions on Image Processing,* vol. 12, no. 9, pp. 1120–1131, Sep. 2003.

[29] B. Lee, S. Hong, H. Lee and E. Kim. "Gait Recognition System using Decision-Level Fusion," in *Proc. 5th IEEE Conference on Industrial Electronics and Applications (ICIEA),* Taichung, Taiwan, pp. 313-316, 2010

[30] I. Vega, and S. Sarkar. "Statistical motion model based on the change of feature relationships: Human gait-based recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence.,* vol. 25, no. 10, pp. 1323–1328, Oct. 2003.

[31] J. Han, and B. Bhanu. "Statistical feature fusion for gait-based human recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Washington, DC, USA, vol. 2, pp. 842–847, 2004

[32] B. Bhanu, and J. Han. "Human recognition on combining kinematic and stationary features," in *Audio- and Video-Based Biometric Person Authentication, LNCS,* vol. 2688, pp. 600-608, 2003

[33] J. Han, and B. Bhanu. "Individual Recognition Using Gait Energy Image," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, no. 2, pp. 316–322, 2006.

[34] Y. Dupuis, X. Savatier, P. Vasseur. "Feature subset selection applied to model-free gait recognition," *Image and Vision Computing Journal*, vol. 31, no. 8, pp. 580-591, 2013

[35] C. Chen, J. Liang, H. Zhao. 'Frame difference energy image for gait recognition with incomplete silhouettes," *Pattern Recognition Letters*, vol. 30, pp. 977-984, 2009

[36]   N. Liu, J. Lu, G. Yang, Y-P Tan. "Robust gait recognition via discriminative set matching," *Journal of Visual Communication and Image Representation*, vol. 24, no. 4, pp. 439-447, 2013

[37]   S. Zheng, J. Zhang, K. Huang, R. He, T. Tan. "Robust view transformation model for gait recognition," in *Proc. 18th IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, pp. 2073-2076, 2011

[38]   X. Ben, W. Meng, R. Yan, K. Wang. "An improved biometrics technique based on metric learning approach," *Neurocomputing Journal*, vol. 97, pp. 44-51, 2012

[39]   Y. Zhang, S. Jiang, Z. Yang, Y. Zhao, and T. Guo. "A Score Level Fusion Framework for Gait-based Human Recognition," in *Proc. 15th IEEE International Workshop on Multimedia Signal Processing* (*MMSP),* Pula, Italy, pp. 189-194, 2013

[40]   P. Thumwarin, K. Srisuk and T. Matsuura. "Gait Identification based on FIR system Characterizing Motion of Leg," in *Proc. 6th Biomedical Engineering International Conference (BMEiCON),* Amphur Muang, Thailand, pp. 1-5, 2013

[41]   K. Moustakas, D. Tzovaras, and G. Stavropoulos. "Gait Recognition Using Geometric Features and Soft Biometrics," *IEEE Signal Processing Letters,* vol. 17, no. 4, pp. 367-370, 2010

[42]   D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, and K. Moustakas. "Gait Recognition Using Compact Feature Extraction Transforms and Depth Information," *IEEE Transactions on Information Forensics and Security,* vol. 2, no. 3, pp. 623-630, Sep. 2007

[43]   D. Gafurov. "A Survey of Biometric Gait Recognition: Approaches, Security and Challenges," in *Proc. NIK conference,* Oslo, Norway, 2007

[44]   L. Liu, W. Jia, and Y. Zhu. "Survey of Gait Recognition," in *Emerging Intelligent Computing Technology and Applications with Aspects of Artificial Intelligence, LNCS*, vol. *5755,* pp. 652–659, 2009

[45]   L. Wang, T. Tan, H. Ning, and W. Hu. "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 14, no. 2, pp. 149–158, Feb. 2004.

[46]   L. Lee, and W. Grimson. "Gait Analysis for Recognition and Classification," in *Proc. Fifth IEEE International Conference on Automatic Face and Gesture Recognition",* Washington, DC, USA, pp. 148–155, 2002

[47]   D. Cunado, M. Nixon, and J. Carter. "Using Gait as a Biometric, via Phase-Weighted Magnitude Spectra," in *Proc. First International Conference on*

Audio- and Video-Based BiometricPerson Authentication (AVBPA), London, UK, pp. 95-102, 1997

[48]   J. Yoo, and M. Nixon. "Markerless Human Gait Analysis via Image Sequences," in *Proc. International Society of Biomechanics XIXth Congress,* Dunedin, New Zealand, 2003

[49]   C. Yam, M. Nixon, and J. Carter. "Automated Person Recognition by Walking and Running via Model-based Approaches," *Pattern Recognition Journal,* vol. 37, no. 5, pp. 1057–1072, 2004

[50]   S. Dockstader, M. Berg, A. Tekalp. "Stochastic Kinematic Modeling and Feature Extraction for Gait Analysis," *IEEE Transactions on Image Processing,* vol. 12, no. 8, pp. 962–976, 2003

[51]   A. F. Bobick, and A. Y. Johnson. "Gait recognition using static, activity-specific parameters," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Kauai, HI, USA, vol. 1, pp. 423-430, 2001

[52]   C. BenAbdelkader, R. Cutler, and L. Davis. "Stride and cadence as a biometric in automatic person identification and verification," in *Proc. Fifth IEEE International Conference on Face and Gesture Recognition,* Washington, DC, USA, *pp. 372–377,* 2002

[53]   R. Tanawongsuwan, and A. Bobick. "Gait recognition from time-normalized joint-angle trajectories in the walking plane," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Kauai, HI, USA, vol. 2, pp. 726–731, 2001

[54]   J. Fortuny-Guasch, P.F. Sammartino, and J. Petit. "Radar techniques for human gait automatic recognition," in *Proc. International Carnahan Conference on Security Technology,* Zurich, Switzerland, *pp. 221-226, 2009*

[55]   Z. Liu, and S. Sarkar. "Simplest representation yet for gait recognition: Averaged silhouette," in *Proc. 17th International Conference on Pattern Recognition,* Cambridge, England, UK, vol. 4, pp. 211–214, 2004

[56]   H. Murase, and R. Sakai. "Moving Object Recognition in Eigenspace Representation: Gait Analysis and Lip Reading," *Pattern Recognition Letters,* vol. 17, no. 2, pp. 155–162, 1996

[57]   P. Huang, C. Harris, and M. Nixon. "Canonical Space Representation for Recognizing Humans by Gait and Face," in *Proc. IEEE Southwest Symposium on Image Analysis and Interpretation,* Tucson, AZ, USA, pp. 180–185, 1998

[58]   H. Su, M. Yang, and H. Xu. "A Novel Method of Gait Recognition Based on Kernel Fisher Discriminant Analysis," in *Proc. IEEE International Conference on Systems, Man and Cybernetics,* Singapore, pp. 830-835, 2008

[59]   J. P. Foster, M. S. Nixon, and A. Prugel-Bennet. "Automatic gait recognition using area-based metrics," *Pattern Recognition Letters,* vol. 24, pp. 2489–2497, 2003.

[60]   J. D. Shutler, and M. S. Nixon. "Zernike velocity moments for description and recognition of moving shapes," in *Proc. British Machine Vision Conference (BMVC),* Manchester, UK, pp. 705–714, 2001

[61]   J. Liu, and N. Zheng. "Gait History Image: A Novel Temporal Template for Gait Recognition," in *Proc. IEEE International Conference on Multimedia and Expo,* Beijing, China, vol. 1-5, pp. 663–666, 2007

[62]   T. Lam, R. Lee, and D. Zhang. "Human gait recognition by the fusion of motion and static spatio-temporal templates," *Pattern Recognition Journal,* vol. 40, pp. 2563-2573, 2007

[63]   T. H.W. Lam, K.H. Cheung, J.K. Liu. "Gait flow image - A silhouette-based gait representation for human identification," *Pattern Recognition Journal,* vol. 44, pp. 973-987, 2011

[64]   E. Hossain, and G. Chetty. "Multimodal Face-Gait Fusion for Biometric Person Authentication," in *Proc. IFIP 9th International Conference on Embedded and Ubiquitous Computing*, Melbourne, Australia, pp. 332-337, 2011

[65]   S. D. Choudhury, and T. Tjahjadi. "Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors," *Pattern Recognition Journal*, vol. 45, pp. 3414-3426, 2012

[66]   M. Jeevan, N. Jain, M. Hanmandlu, and G. Chetty. "Gait Recognition Based on Gait Pal and Pal Entropy Image," in *Proc. 20th IEEE International Conference on Image Processing (ICIP),* Melbourne, Australia, pp. 4195-4199, 2013

[67]   C. P. Lee, A. W. Tan, and S. C. Tan. "Time-sliced averaged motion history image for gait recognition," *Journal of Visual Communication & Image Representation*, vol. 25, pp. 822-826, 2014

[68]   A. Jain, and A. Ross. "Multibiometric systems," in *Communications of the ACM - Multimodal interfaces that flex, adapt, and persist*, vol. 47, no. 1, pp. 34-40, Jan 2004.

[69]   A. Ross, and A. K. Jain. "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, pp. 2115-2125, Sep 2003.

[70]   S. Zheng, K. Huang, T. Tan, and D. Tao. "A cascade fusion scheme for gait and cumulative foot pressure image recognition," *Pattern Recognition Journal*, vol. 45, no. 10, pp. 3603-3610, 2012.

[71]   R. Chellapa, A. K. Roy-Chowdhury, and A. Kale. "Human Identification using Gait and Face," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Minneapolis, MN, USA, pp. 1-2, 2007

[72]   L. Wang, H. Ning, T. tan, and W. Hu. "Fusion of Static and Dynamic Body Biometrics for Gait Recognition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 14, no. 2, pp. 149-158, 2003

[73]   L. Lam, and C. Suen. "Application of majority voting to pattern recognition: an analysis of its behavior and performance," *IEEE Transactions on System, Man and Cybernetics*, vol. 27, no. 5, pp. 553-568, 1997

[74]   L. Xu, A. Krzyak, and C.Y. Suen. "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on System, Man and Cybernetics*, vol. 22, no. 3, pp. 418-435, 1992

[75]   R. Maclin, and J. Shavlik. "Combining the Prediction of Multiple Classifiers: Using Competitive Learning to Initialize Neural Networks," in *Proc. 14th International Joint Conference on Artificial Intelligence*, Montreal, Quebec, Canada, 1995

[76]   G. Rogova, and R. Menon. "Decision Fusion for Learning in Pattern Recognition," in *Proc. FUSION'98, First International Conference on Multisource-Multisensor Information Fusion,* Las Vegas, Nevada, USA, 1998

[77]   G. Rogova. "The Dempster-Shafer theory of evidence for classification of medical images," in *Proc. the 6th IEEE Dual-Use Technologies and Application Conference*, Syracuse, NY, USA, 1996

[78]   J. Yang, D. Zhang, A. F. Frangi, and J. Yang. "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131-137, 2004

[79]   J. Ye. "Generalized low rank approximations of matrices," in *Proc. 21st International Conference on Machine Learning*, Alberta, Canada, pp. 887-894, 2004

[80]   D. Xu, S. Yan, L. Zhang, S. Lin, H.-J. Zhang, and T. S. Huang. "Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis," *IEEE Transactions on Circuits and Systems for Video Technology*. vol. 18, no. 1, pp. 36-47, 2008

[81]   J. Ye, R. Janardan, Q. Li. "GPCA: An efficient dimension reduction scheme for image compression and retrieval," in *Proc. The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, Washington, USA, pp. 354-363, 2004.

[82] D. Tao, M. Song, X. Li, J. Shen, J. Sun, X. Wu, C. Faloutsos, and S. Maybank. "Bayesian tensor approach for 3-D face modeling," *IEEE Transactions on Circuits and Systems for Video Technology*. vol. 18, no. 10, pp. 1397-1410, 2008.

[83] K. Inoue, K. Hara, and K. Urahama. "Robust multilinear principal component analysis," in *Proc. IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, pp. 591-597, 2009

[84] Y. Panagakis, C. Kotropoulos, and G. Arce. "Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 576-588, 2010

[85] J. Sun, D. Tao, S. Papadimitriou, P. Yu, and C. Faloutsos. "Incremental tensor analysis: Theory and applications," *ACM Transactions on Knowledge Discovery from Data*, vol. 2, no. 3, pp. 11:1-11:37, 2008

[86] A. Shashua, and A. Levin. "Linear image coding for regression and classification using the tensor-rank principle," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 42-49, 2001

[87] H. Lu, K. Plataniotis, and A. Venetsanopoulos. "Uncorrelated multilinear principal component analysis for unsupervised multilinear subspace learning," *IEEE Transactions on Neural Networks*, vol. 20, no. 11, pp. 1820-1836, 2009

[88] J. Ye, R. Janardan, and Q. Li. "Two-dimensional linear discriminant analysis," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, Vancouver and Whistler, British Columbia, Canada, pp. 1569-1576, 2004.

[89] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang. "Multilinear discriminant analysis for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 212-220, 2007.

[90] D. Tao, X. Li, X. Wu, and S. Maybank. "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no.10, pp. 1700-1715, 2007.

[91] Y. Wang, and S. Gong. "Tensor discriminant analysis for view-based object recognition," in *Proc. 18th International Conference on Pattern Recognition*, Hong Kong, vol. 3, pp. 33-36, 2006.

[92] H. Lu, K. Plataniotis, and A. Venetsanopoulos. "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 103-123, 2009

[93]   J. Han, and B. Bhanu. "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316-322, Feb. 2006

[94]   E. Zhang, H. Ma, J. Lu, and Y. Chen. "Gait Recognition using Dynamic Gait Energy and PCA+LPP Method," in *Proc. 8th International Conference on Machine Learning and Cybernetics*, Baoding, China, vol. 1, pp. 50-53, 12-15 Jul, 2009

[95]   Y. Pratheepan, J. Condell, and G. Prasad. "The use of dynamic and Static Characteristics of gait for individual identification," in *Proc. 13th International Machine Vision and Image Processing Conference*, Dublin, Ireland, pp. 111-116, 2009

[96]   L. Wang, S. Jia, X. Li, and S. Wang. "Human Gait Recognition Based on Gait Flow Image Considering Walking Direction," in *Proc. IEEE International Conference on Mechatronics and Automation*, Chengdu, China, pp. 1990-1995, 5-8 Aug, 2012

[97]   P. O'Donovan. "Optical Flow Techniques and Applications," *The University of Saskatchewan*, Apr 6, 2005

[98]   B. Horn, and B. Schunck. "Determining Optical Flow," *Artificial Intelligence Journal*, vol. 17, no. 1-3, pp. 185-203, 1981

Vita

Amer Al-Tayyan was born on February 6, 1979, in Irbid, Jordan. He was educated in public schools in Abu Dhabi, UAE and graduated with distinction in 1997. Soon after graduation from high school, Amer joined Ajman University of Science and Technology, from which he graduated with honor in 2003, receiving Bachelor of Science in Electrical Engineering.

Amer has, since then, enrolled in various internship and training programs in industrial automation, computer hardware and software, medical equipment, project management and security systems. It was the field of security systems in which he has become experienced for the past 10 years, moving from being Project Engineer to Senior Engineer at Citytec (CG Group) in Dubai, and holding a current position of Security Consultant at WSP, a global engineering consulting firm, with the Middle East office located in Sharjah.

In 2009, Amer began a Master's program in Electrical Engineering at the American University of Sharjah. His research interests include computer vision, pattern recognition, biometrics and image and video processing.

Amer is a member of the IEEE, IEEE Computational Intelligence Society and ASIS International.